A comparison of the Extrapolated Successive Overrelaxation and the Preconditioned Simultaneous Displacement methods for augmented linear systems

M. A. Louka · N. M. Missirlis

Abstract In this paper we study the impact of two types of preconditioning on the numerical solution of large sparse augmented linear systems. The first preconditioning matrix is the lower triangular part whereas the second is the product of the lower triangular part with the upper triangular part of the augmented system's coefficient matrix. For the first preconditioning matrix we form the Generalized Modified Extrapolated Successive Overrelaxation (GMESOR) method, whereas the second preconditioning matrix yields the Generalized Modified Preconditioned Simultaneous Displacement (GMPSD) method, which is an extrapolated form of the Symmetric Successive Overrelaxation method. We find sufficient conditions for each aforementioned iterative method to converge. In addition, we develop a geometric approach, for determining the optimum values of their parameters and corresponding spectral radii. It is shown that both iterative methods studied (GMESOR and GMPSD) attain the same rate of convergence. Numerical results confirm our theoretical expectations.

M. A. Louka · N. M. Missirlis
Department of Informatics and Telecommunications, University of Athens,
Panepistimiopolis, 15784, Athens, Greece
Tel.: +30-72-75103
Fax: +30-72-75114
E-mail: mlouka@di.uoa.gr; nmis@di.uoa.gr

1 Introduction

Let $A \in \mathbb{R}^{m \times m}$ be a symmetric positive definite matrix and $B \in \mathbb{R}^{m \times n}$ be a matrix of full column rank, where $m \geq n$. Then, the augmented linear system is of the form

$$\mathcal{A}u = b \tag{1}$$

where

$$\mathcal{A} = \begin{pmatrix} A & B \\ -B^T & 0 \end{pmatrix}, \ \ u = \begin{pmatrix} x \\ y \end{pmatrix}, \ \ b = \begin{pmatrix} b_1 \\ -b_2 \end{pmatrix} \tag{2}$$

with B^T denoting the transpose of the matrix B.

When A and B are large and sparse matrices, iterative methods for solving (1)-(2) are effective and more attractive than direct methods, because of storage requirements and preservation of sparsity. There are several approaches to the iterative solution of (1)-(2). First, we mention multigrid methods [13], [35], which are often the most efficient and effective methods for solving large, sparse, linear systems [15], [48]. For example, one can apply multigrid techniques to the whole system (1)-(2) to solve problems in areas of computational fluid dynamics [22], [30], [40], [51], [53], [24], [25] constrained optimization [43], [44], [45], [46], mixed finite elements [1], [23] and elsewhere. For parallel multigrid see e.g [26], [27], [28].

On the other hand the difficulty in applying iterative methods such as the Successive Overrelaxation (SOR) method [55] to the system (1)-(2) is the singularity of the block diagonal part of the coefficient matrix. Various methods have been developed to overcome this problem such as the Uzawa and the Preconditioned Uzawa methods [2], [14], [16]. In 2001, Golub et al. [21] generalized the Uzawa and the Preconditioned Uzawa methods by introducing an additional acceleration parameter and produced the SOR-like method. When a good preconditioning matrix is easily computed one can consider the MINRES and GMRES methods [20], [14] for solving (1)-(2). In case the matrix A in (2) is symmetric and positive definite, the Preconditioned Conjugate Gradient (PCG) method [31] can be applied. This was done with an SOR-like preconditioner in the work by Li, Evans and Zhang in [33]. In 2005, Bai et al. [8] studied the Generalized SOR (GSOR) method by introducing an additional parameter to the SOR-like method and proved that it possesses the same rate of convergence but lower complexity than the PCG method. Furthermore, the Generalized Modified Extrapolated SOR (GMESOR) method was also proposed for further study. The latter is a generalization of the GSOR method as it uses one additional parameter. The way of introducing parameters resembles the one followed for the formulation of the Modified SOR method [55], [29], [36], [37], [38] in case of two-cyclic linear systems.

The present paper was motivated by the work in [8]. We develop the convergence analysis of the Generalized Modified Extrapolated SOR (GMESOR) method and the Generalized Modified Preconditioned Simultaneous Displacement (GMPSD) method. These methods introduce more parameters with the

hope to further increase their rate of convergence. The goal of our work was to study the impact of two different preconditioning matrices to the convergence rate of the associated iterative method for solving the augmented linear system (1)-(2). First, we use the preconditioning matrix which is formed by the lower triangular part of $\mathcal{A}$ and formulate the GMESOR method which is an extrapolated form of the GSOR method. Secondly, we consider as preconditioning matrix the product of the lower with the upper triangular part of $\mathcal{A}$ and construct the GMPSD method. The reason for studying the latter form of preconditioning matrix is to obtain a better approximation to the matrix $\mathcal{A}$ than the former and as such it is hoped to produce an iterative method with a faster rate of convergence. The construction of both methods resembles the one followed for the MESOR and MPSD methods studied in [37] and [38], respectively, for two-cyclic matrices. Our starting point, for studying these iterative methods, is the derivation of functional relationships which relate the eigenvalues of their iteration matrices with those of the matrix $J = Q^{-1}B^T A^{-1}B$. Assuming that the matrix Q is symmetric positive or negative definite, the eigenvalues of the matrix J are real and either positive or negative, respectively. Under these assumptions we find sufficient conditions for the convergence of the GMESOR and GMPSD methods and determine the optimum values of their parameters. The study of GMESOR and GMPSD becomes interesting as these methods can be used either as preconditioners to Krylov subspace methods [7], [21], [31], [20] or as smoothers to multilevel methods [3], [4], [5]. Traditionally, multigrid methods utilize stationary iterative methods (such as Jacobi, Gauss-Seidel) to smooth out high-frequency errors and accelerate the convergence. In [32] a semi-iterative method, namely the Chebyshev-Jacobi method, was used as smoother. Similarly, the GMESOR method or the GMPSD method in combination with semi-iterative techniques can be used as smoothers to accelerate the rate of convergence of multigrid methods. Recent work for the application of algebraic multigrids for saddle point systems is presented in [35] and the references therein.

The contributions of our paper can be summarized as follows.
(i) We present a simple and unified approach for developing the convergence analysis of the GMESOR and GMPSD methods. In particular, we develop a geometrical approach for the determination of the optimum values of the parameters in GMESOR and GMPSD methods which is similar to Varga [50] p. 111, for finding the optimum value of the parameter ω in SOR. The difference, in our case, is that now the functional relationship contains more than two parameters and consequently we had to extend the proof of [50]. There is a different algebraic approach in [55] pp. 279 for the determination of the two optimum values for ω and $\omega^{'}$ in the Modified SOR (MSOR) method which, with some additional modifications, will solve the problem as far as the GSOR method is concerned. Nevertheless, it is doubtful whether this approach works also for the determination of the optimum value for more than two parameters as is the case for the GMESOR and GMPSD methods. This is also the case if one adopts the approach of [8].
(ii) From our theoretical and experimental analysis it is shown that both afore-

mentioned forms of preconditioning matrices have the same impact on the convergence rate of the induced iterative method for the numerical solution of the augmented linear system (1)-(2). More specifically, the GMESOR and GMPSD methods attain the same convergence rate since their spectral radii are identical for the optimum values of their parameters. In particular, we show that GMESOR degenerates to GSOR, whereas a simplified version of the GMPSD method is identical to a backward form of the GSOR method. Furthermore, we compare the effectiveness of our methods in relation to the PHSS [6], [7], [9], [10], [11], [52] and Krylov subspace methods [41], [42], [49]. The paper is organized as follows. In section 2 we study the convergence of the GMESOR method. In particular, we find sufficient conditions for GMESOR to converge under the assumption that the eigenvalues of the J matrix are real. We also determine optimum values for its parameters. A similar convergence analysis for the GMPSD method is developed in section 3. In section 4, we present our numerical results and finally in section 5 we state our remarks and conclusions.

2 The Generalized Modified Extrapolated SOR (GMESOR) method

Let the coefficient matrix $\mathcal{A}$ of (1) be defined by the splitting

$$\mathcal{A} = \mathcal{D} - \mathcal{L} - \mathcal{U} \tag{3}$$

where

$$\mathcal{D} = \begin{pmatrix} A & 0 \\ 0 & Q \end{pmatrix}, \ \mathcal{L} = \begin{pmatrix} 0 & 0 \\ B^T & aQ \end{pmatrix}, \ \mathcal{U} = \begin{pmatrix} 0 & -B \\ 0 & (1-a)Q \end{pmatrix}, \tag{4}$$

with $Q \in \mathbb{R}^{n \times n}$ be a prescribed nonsingular and symmetric matrix and $a \in \mathbb{R}$. Furthermore, we denote by T, the diagonal matrix $T = diag(\tau_1 I_m, \tau_2 I_n)$ with $\tau_1, \tau_2 \in \mathbb{R} - \{0\}$, $I_m \in \mathbb{R}^{m \times m}$ and $I_n \in \mathbb{R}^{n \times n}$ be identity matrices.
For the numerical solution of (1), we consider the following iterative scheme

$$\begin{pmatrix} x^{(k+1)} \\ y^{(k+1)} \end{pmatrix} = \mathcal{H}(\tau_1, \tau_2) \begin{pmatrix} x^{(k)} \\ y^{(k)} \end{pmatrix} + \eta(\tau_1, \tau_2) \begin{pmatrix} b_1 \\ -b_2 \end{pmatrix} \tag{5}$$

where

$$\mathcal{H}(\tau_1, \tau_2) = I - R^{-1} T \mathcal{A}, \ \ \eta(\tau_1, \tau_2) = R^{-1} T b, \tag{6}$$

R is a nonsingular matrix to be defined and $I = diag(I_m, I_n)$.
In the sequel we consider two different types of preconditioning matrices R and study the corresponding iterative methods derived by (5) and (6).

2.1 The functional relationship

As a first step we consider the preconditioning matrix which is formed by the parameterized diagonal and lower triangular part of $\mathcal{A}$

$$R = \mathcal{D} - \Omega\mathcal{L}, \tag{7}$$

where $\Omega = diag(\omega_1 I_m, \omega_2 I_n)$ with $\omega_1, \omega_2 \in \mathbb{R}$. If R is given by (7), then (6) becomes

$$\mathcal{H}(\tau_1, \tau_2, \omega_2, a) = I - (\mathcal{D} - \Omega\mathcal{L})^{-1}T\mathcal{A}$$

or

$$\mathcal{H}(\tau_1, \tau_2, \omega_2, a) = (\mathcal{D} - \Omega\mathcal{L})^{-1}[(I - T)\mathcal{D} + (T - \Omega)\mathcal{L} + T\mathcal{U}] \tag{8}$$

and

$$\eta(\tau_1, \tau_2, \omega_2, a) = (\mathcal{D} - \Omega\mathcal{L})^{-1}Tb. \tag{9}$$

Note that the parameter ω_1 is absent in $\mathcal{H}$ and η. This is because the first m rows of $\mathcal{L}$ are zeros a fact which is carried over in matrix $\Omega\mathcal{L}$ also.

The iterative scheme given by (5),(6),(8) and (9) will be referred to as the Generalized Modified Extrapolated SOR (GMESOR) method. In case $a = 0$ this method was introduced in [8] and proposed for further study. In the sequel to distinguish the dependence of GMESOR upon the parameter a we use the notation GMESOR(a).

For $(\mathcal{D} - \Omega\mathcal{L})^{-1}$ to exist we require

$$\det(\mathcal{D} - \Omega\mathcal{L}) \neq 0.$$

Because of (4)

$$R = \mathcal{D} - \Omega\mathcal{L} = \begin{pmatrix} A & 0 \\ -\omega_2 B^T & (1 - a\omega_2)Q \end{pmatrix}.$$

Therefore,

$$\det(\mathcal{D} - \Omega\mathcal{L}) = (1 - a\omega_2)^n \det(A) \det(Q) \neq 0$$

or

$$a\omega_2 \neq 1 \tag{10}$$

since the matrix A is symmetric positive definite and the matrix Q is nonsingular. In the sequel we require (10) to hold.

The GMESOR(a) method has the following algorithmic form.

THE GMESOR(a) METHOD: *Let $Q \in \mathbb{R}^{n \times n}$ be a nonsingular and symmetric matrix. Given initial vectors $x^{(0)} \in \mathbb{R}^m$ and $y^{(0)} \in \mathbb{R}^n$, and the parameters $\tau_1, \tau_2 \neq 0$, $\omega_2, a \in \mathbb{R}$ with $a\omega_2 \neq 1$. For $k = 0, 1, 2, ...$ until the iteration sequence $\{(x^{(k)^T}, y^{(k)^T})^T\}$ is convergent, compute*

$$x^{(k+1)} = (1 - \tau_1)x^{(k)} + \tau_1 A^{-1}(b_1 - By^{(k)}),$$

$$y^{(k+1)} = y^{(k)} + \frac{1}{1 - a\omega_2} Q^{-1} \left\{ B^T[\omega_2 x^{(k+1)} + (\tau_2 - \omega_2)x^{(k)}] - \tau_2 b_2 \right\},$$

where Q is an approximation of the Schur complement matrix $B^T A^{-1} B$.

For special values of its parameters GMESOR(a) degenerates into known methods. Indeed, if $\omega = \tau_1 = \tau_2 = \omega_2$ and $a = 0$ then GMESOR becomes the SOR-like method [21]; if $\omega = \tau_1 = \tau_2 = \omega_2 = 1$ and $a = 0$ then it becomes the preconditioned Uzawa method [16]; and if $\tau_1 = \omega_1$, $\tau_2 = \omega_2$ and $a = 0$, then it becomes the GSOR method [8]. By comparing the algorithmic structures of GMESOR(a) and GSOR, one can verify that the former has an additional matrix times a vector computation. Finally, if

$$\tau_1 = \omega, \quad \frac{\omega_2}{1 - a\omega_2} = \gamma \ \text{ and } \ \frac{\tau_2}{1 - a\omega_2} = \tau \tag{11}$$

then the GMESOR(a) method becomes the Generalized Inexact Accelerated Overrelaxation (GIAOR) method [8] and if

$$\tau_1 = \omega, \tau_2 = \omega_2 \text{ and } \ \frac{\tau_2}{1 - a\omega_2} = \tau \tag{12}$$

the GMESOR(a) method becomes the Parametrized Inexact Uzawa (PIU) method [12] when $P = A$. The following theorem establishes the functional relationship between the eigenvalues λ of the iteration matrix $\mathcal{H}(\tau_1, \tau_2, \omega_2, a)$ and the eigenvalues μ of the associated matrix $J = Q^{-1} B^T A^{-1} B$.

Theorem 2.1 *Let $A \in \mathbb{R}^{m \times m}$ be symmetric positive definite, $B \in \mathbb{R}^{m \times n}$ be of full column rank and $Q \in \mathbb{R}^{n \times n}$ be nonsingular and symmetric. If $\lambda \neq 1 - \tau_1$ is an eigenvalue of the matrix $\mathcal{H}(\tau_1, \tau_2, \omega_2, a)$ of the GMESOR(a) method and if μ satisfies*

$$\lambda^2 + \lambda \left(\tau_1 - 2 + \frac{\tau_1 \omega_2}{1 - a\omega_2} \mu \right) + 1 - \tau_1 + \frac{\tau_1(\tau_2 - \omega_2)}{1 - a\omega_2} \mu = 0, \tag{13}$$

where $a\omega_2 \neq 1$, then μ is an eigenvalue of the matrix $J = Q^{-1} B^T A^{-1} B$. Conversely, if μ is an eigenvalue of J and if $\lambda \neq 1 - \tau_1$ satisfies (13), then λ is an eigenvalue of $\mathcal{H}(\tau_1, \tau_2, \omega_2, a)$. In addition, $\lambda = 1 - \tau_1$ is an eigenvalue of $\mathcal{H}(\tau_1, \tau_2, \omega_2, a)$ (if $m > n$) with the corresponding eigenvector $(x^T, 0)^T$, where $x \in \mathcal{N}(B^T)$ and $\mathcal{N}(B^T)$ is the null space of B^T.

Proof The eigenvalues μ of the matrix $J = Q^{-1} B^T A^{-1} B$ are real, positive and non-zero. Let λ be a nonzero eigenvalue of the iteration matrix $\mathcal{H}(\tau_1, \tau_2, \omega_2, a)$ defined in (8), and $(x, y)^T \in \mathbb{R}^{m+n}$ be the corresponding eigenvector. Then,

$$\mathcal{H}(\tau_1, \tau_2, \omega_2, a) \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix} \tag{14}$$

or, from (8) we have

$$[(I - T)\mathcal{D} + (T - \Omega)\mathcal{L} + T\mathcal{U}] \begin{pmatrix} x \\ y \end{pmatrix} = \lambda (\mathcal{D} - \Omega\mathcal{L}) \begin{pmatrix} x \\ y \end{pmatrix}. \tag{15}$$

From (124) and (4) it follows that

$$\begin{pmatrix} (1-\tau_1)A & -\tau_1 B \\ (\tau_2 - \omega_2)B^T & (1-a\omega_2)Q \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} A & 0 \\ -\omega_2 B^T & (1-a\omega_2)Q \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Decoupling we obtain

$$\begin{cases} (1-\tau_1)Ax - \tau_1 By = \lambda A x \\ (\tau_2 - \omega_2)B^T x + (1-a\omega_2)Qy = -\lambda\omega_2 B^T x + \lambda(1-a\omega_2)Qy \end{cases}$$

or equivalently

$$\begin{cases} (1-\tau_1 - \lambda)x = \tau_1 A^{-1}By \\ (1-\lambda)(1-a\omega_2)y = [(1-\lambda)\omega_2 - \tau_2]Q^{-1}B^T x. \end{cases} \tag{16}$$

Multiplying the first equality in (125) by $Q^{-1}B^T$, we obtain

$$(1-\tau_1 - \lambda)Q^{-1}B^T x = \tau_1 Q^{-1}B^T A^{-1}By,$$

or, when $\lambda \neq 1 - \tau_1$, we have

$$Q^{-1}B^T x = \frac{\tau_1}{1 - \tau_1 - \lambda}Q^{-1}B^T A^{-1}By. \tag{17}$$

From (126) and the second equality in (125) it follows that

$$(1-\lambda)(1-a\omega_2)(1-\lambda-\tau_1)y = [(1-\lambda)\omega_2 - \tau_2]\tau_1 Jy. \tag{18}$$

If $\lambda = 1 - \tau_1 \neq 0$, we have from (125) that $By = 0$ and $\tau_1(1-a\omega_2)Qy = (\tau_1\omega_2 - \tau_2)B^T x$. It then follows that $y = 0$ and $x \in \mathcal{N}(B^T)$. Hence, $\lambda = 1 - \tau_1$ is an eigenvalue of $\mathcal{H}(\tau_1, \tau_2, \omega_2, a)$ with the corresponding eigenvector $(x^T, 0)^T$, where $x \in \mathcal{N}(B^T)$. Therefore, because of (18), the eigenvalues λ (except for $\lambda = 1 - \tau_1$) of the iteration matrix $\mathcal{H}(\tau_1, \tau_2, \omega_2, a)$ of the GMESOR(a) method and the eigenvalues μ of the matrix J are related through the functional relationship

$$(1-\lambda)(1-a\omega_2)(1-\lambda-\tau_1) = [(1-\lambda)\omega_2 - \tau_2]\tau_1\mu,$$

namely λ satisfies the quadratic equation (13). $\qquad\qquad\square$

From the above theorem we can obtain the following functional relationships for the GESOR(a), SOR-like(a) and GSOR(a) methods.

Corollary 2.1 *Under the hypothesis of Theorem 2.1*

1. The nonzero eigenvalues of the iteration matrix $\mathcal{H}(\tau, \omega_2, a)$ of the GESOR(a) method are given by $\lambda = 1 - \tau$ or if $a\omega_2 \neq 1$ by

$$\lambda^2 + \lambda\left(\tau - 2 + \frac{\tau\omega_2}{1 - a\omega_2}\mu\right) + 1 - \tau + \frac{\tau(\tau - \omega_2)}{1 - a\omega_2}\mu = 0. \tag{19}$$

2. The nonzero eigenvalues of the iteration matrix $\mathcal{L}(\omega, a)$ of the SOR-like(a) method are given by $\lambda = 1 - \omega$ or if $a\omega \neq 1$ by

$$\lambda^2 + \lambda\left(\omega - 2 + \frac{\omega^2}{1 - a\omega}\mu\right) + 1 - \omega = 0. \tag{20}$$

3. The nonzero eigenvalues of the iteration matrix $\mathcal{L}(\omega_1, \omega_2, a)$ of the GSOR(a) method are given by $\lambda = 1 - \omega_1$ or if $a\omega_2 \neq 1$ by

$$\lambda^2 + \lambda \left(\omega_1 - 2 + \frac{\omega_1 \omega_2}{1 - a\omega_2} \mu \right) + 1 - \omega_1 = 0. \tag{21}$$

Proof The iteration matrix $\mathcal{H}(\tau, \omega_2, a)$ is obtained by letting $\tau = \tau_1 = \tau_2$ in $\mathcal{H}(\tau_1, \tau_2, \omega_2, a)$. By following a similar approach as in the proof of Theorem 2.1 we find the functional relationship (19). Similarly, we find (20) and (21). □ Note that the above functional relationships are generalizations of the original SOR-like and GSOR methods. Indeed, if $a = 0$, then from (20) we obtain the functional relationship of the SOR-like method [21], whereas from (21) we obtain the functional relationship of the GSOR method [8]. □

Another preconditioning matrix R, which is formed by the upper triangular part of $\mathcal{A}$ is the following

$$R = \mathcal{D} - \Omega \mathcal{U}. \tag{22}$$

Using (22) in (6) then (5) becomes the backward form of the GMESOR(a) method, which will be referred to as the Generalized Modified Extrapolated Backward SOR(a) (GMEBSOR(a)) method. From (6), because of (22), the iteration matrix of the GMEBSOR(a) method is given by

$$\mathcal{K}(\tau_1, \tau_2, \omega_1, \omega_2, a) = I - (\mathcal{D} - \Omega \mathcal{U})^{-1} T \mathcal{A}$$

or

$$\mathcal{K}(\tau_1, \tau_2, \omega_1, \omega_2, a) = (\mathcal{D} - \Omega \mathcal{U})^{-1} [(I - T)\mathcal{D} + (T - \Omega)\mathcal{U} + T\mathcal{L}] \tag{23}$$

and

$$k(\tau_1, \tau_2, \omega_1, \omega_2, a) = (\mathcal{D} - \Omega \mathcal{U})^{-1} T b. \tag{24}$$

For $(\mathcal{D} - \Omega \mathcal{U})^{-1}$ to exist we require

$$\det(\mathcal{D} - \Omega \mathcal{U}) \neq 0. \tag{25}$$

Because of (4)

$$R = \mathcal{D} - \Omega \mathcal{U} = \begin{pmatrix} A & -\omega_1 B \\ 0 & [1 - (1 - a)\omega_2]Q \end{pmatrix}. \tag{26}$$

Therefore, (25) becomes

$$\det(\mathcal{D} - \Omega \mathcal{U}) = [1 - (1 - a)\omega_2]^n \det A \det Q \neq 0$$

or

$$(1 - a)\omega_2 \neq 1 \tag{27}$$

since the matrix A is symmetric positive definite and the matrix Q is nonsingular. The GMEBSOR(a) method has the following algorithmic form.

THE GMEBSOR(a) METHOD: *Let $Q \in \mathbb{R}^{n \times n}$ be a nonsingular and symmetric matrix. Given initial vectors $x^{(0)} \in \mathbb{R}^m$ and $y^{(0)} \in \mathbb{R}^n$, and the parameters $\tau_1, \tau_2 \neq 0$, $\omega_1, \omega_2, a \in \mathbb{R}$ with $(1 - a)\omega_2 \neq 1$. For $k = 0, 1, 2, \ldots$ until the*

iteration sequence $\{(x^{(k)^T}, y^{(k)^T})^T\}$ *is convergent, compute*

$$y^{(k+1)} = y^{(k)} + \frac{\tau_2}{1 - (1-a)\omega_2} Q^{-1}(B^T x^{(k)} - b_2)$$

$$x^{(k+1)} = (1 - \tau_1)x^{(k)} + A^{-1}\left\{\tau_1(b_1 - By^{(k)}) - \omega_1 B(y^{(k+1)} - y^{(k)})\right\},$$

where Q is an approximate (preconditioning) matrix of the Schur complement matrix $B^T A^{-1} B$.

As a by-product of the GMEBSOR(a) method we obtain the backward schemes corresponding to the GESOR(a) and GSOR(a) methods. For $\tau = \tau_1 = \tau_2$, we have the GEBSOR(a) method and for $\tau_1 = \omega_1$ and $\tau_2 = \omega_2$ we have the GBSOR(a) method.

Corollary 2.2 *Under the hypothesis of Theorem 2.1*
1. The nonzero eigenvalues of the iteration matrix $\mathcal{K}(\tau_1, \tau_2, \omega_1, \omega_2, a)$ of the GMEBSOR(a) method are given by $\lambda = 1 - \tau_1$ or if $(1-a)\omega_2 \neq 1$ by

$$\lambda^2 + \lambda\left(\tau_1 - 2 + \frac{\tau_2\omega_1}{1 - (1-a)\omega_2}\mu\right) + 1 - \tau_1 + \frac{\tau_2(\tau_1 - \omega_1)}{1 - (1-a)\omega_2}\mu = 0. \quad (28)$$

2. The nonzero eigenvalues of the iteration matrix $\mathcal{K}(\tau, \omega_1, \omega_2, a)$ of the GEBSOR(a) method are given by $\lambda = 1 - \tau$ or if $(1-a)\omega_2 \neq 1$ by

$$\lambda^2 + \lambda\left(\tau - 2 + \frac{\tau\omega_1}{1 - (1-a)\omega_2}\mu\right) + 1 - \tau + \frac{\tau(\tau - \omega_1)}{1 - (1-a)\omega_2}\mu = 0. \quad (29)$$

3. The nonzero eigenvalues of the iteration matrix $\mathcal{M}(\omega_1, \omega_2, a)$ of the GBSOR(a) method are given by $\lambda = 1 - \omega_1$ or if $(1-a)\omega_2 \neq 1$ by

$$\lambda^2 + \lambda\left(\omega_1 - 2 + \frac{\omega_1\omega_2}{1 - (1-a)\omega_2}\mu\right) + 1 - \omega_1 = 0. \quad (30)$$

Proof Following a similar approach as in the proof of Theorem 2.1 and using the iteration matrix $\mathcal{K}(\tau_1, \tau_2, \omega_1, \omega_2, a)$ given by (23), we find the functional relationship (28). Similarly, we find the functional relationships (29) and (30). □

Note that the GMEBSOR(a) method has four parameters instead of three as the GMESOR(a) method whereas the GEBSOR(a) method has three parameters instead of two as the GESOR(a) method. If $a = 1$, then (30) becomes the functional relationship of the GSOR(a) method.

2.2 Convergence

In this section we develop the convergence analysis of the GSOR(a) and GMESOR method s as well as their corresponding backward counterparts. In particular, we derive sufficient conditions for the GSOR(a) and the GMESOR method to converge under the assumption that the eigenvalues of the matrix

J are all real. The sign of J's eigenvalues depends upon the properties of the matrix Q. Specifically, we assume that the matrix Q is symmetric positive or negative definite.

2.2.1 The GSOR(a) method

The next theorem provides sufficient conditions for the GSOR(a) method to converge if the matrix Q is symmetric positive definite and $a \neq 0$.

Theorem 2.2 *Let $A \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m \times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1}B$ by μ_{min} and μ_{max}, respectively. Then $\rho(\mathcal{L}(\omega_1, \omega_2, a)) < 1$ if the parameters ω_1 and ω_2 lie in any case of Table 1.*

Table 1 Sufficient conditions for the GSOR(a) method to converge if $\mu_{min} > 0$.

Condition	Cases	$\omega_2 -$ Domain	$\omega_1 -$ Domain
$a > 0$	1	$0 < \omega_2 < \dfrac{2(2-\omega_1)}{\omega_1\mu_{max} + 2a(2-\omega_1)}$	$0 < \omega_1 < 2$
$a < 0$	2	$\omega_2 < \dfrac{2(2-\omega_1)}{\omega_1\mu_{max} + 2a(2-\omega_1)} (< 0)$	$0 < \omega_1 < \dfrac{4a}{2a - \mu_{max}}$
	3	$0 < \omega_2$	$0 < \omega_1 < \dfrac{4a}{2a - \mu_{max}}$
	4	$0 < \omega_2 < \dfrac{2(2-\omega_1)}{\omega_1\mu_{max} + 2a(2-\omega_1)}$	$\dfrac{4a}{2a - \mu_{min}} < \omega_1 < 2$

Proof Recall (Corollary 2.1) that $\lambda = 1-\omega_1 \neq 0$ is an eigenvalue of $\mathcal{L}(\omega_1, \omega_2, a)$ and if $\lambda \neq 1 - \omega_1$ then the eigenvalues of $\mathcal{L}(\omega_1, \omega_2, a)$ are given by (21). If $\lambda = 1-\omega_1 \neq 0$, then the GSOR($a$) method is convergent if and only if $|\lambda| < 1$, that is $|1 - \omega_1| < 1$, or

$$0 < \omega_1 < 2. \tag{31}$$

If $\lambda \neq 1 - \omega_1$ and $a\omega_2 \neq 1$, then (21) holds and by Lemma 2.1 page 171 of [55] it follows that the GSOR(a) method is convergent if and only if

$$|c| < 1 \quad \text{and} \quad |b| < 1 + c \tag{32}$$

where

$$c = 1 - \omega_1 \tag{33}$$

and

$$b = 2 - \omega_1 - \frac{\omega_1\omega_2\mu}{1 - a\omega_2}. \tag{34}$$

From the first part of (136), because of (33), it follows that (31) holds also in this case. From the second part of (136), because of (33) and (34), it follows that

$$\left| 2 - \omega_1 - \frac{\omega_1\omega_2\mu}{1 - a\omega_2} \right| < 2 - \omega_1$$

or

$$0 < \frac{\omega_2}{1 - a\omega_2} < \frac{2(2 - \omega_1)}{\omega_1 \mu}. \tag{35}$$

In order for (35) to hold we distinguish two cases. Case I: $\omega_2 > 0$ and $1 - a\omega_2 > 0$ and Case II: $\omega_2 < 0$ and $1 - a\omega_2 < 0$. For each of theses cases we will distinguish two subcases. (i) $a > 0$ (ii) $a < 0$. In the sequel we will study the subcase (i) of Case I, since the other cases can be treated similarly. For subcase (i) of Case I

$$0 < \omega_2 < \frac{1}{a}. \tag{36}$$

From (35), we have

$$[\omega_1 \mu + 2a(2 - \omega_1)]\omega_2 < 2(2 - \omega_1). \tag{37}$$

We distinguish two subcases: (i$_1$) $\omega_1 \mu + 2a(2 - \omega_1) > 0$ and (i$_2$) $\omega_1 \mu + 2a(2 - \omega_1) < 0$. In the sequel we will only treat subcase (i$_1$) since the other case can be treated similarly. If $\omega_1 \mu + 2a(2 - \omega_1) > 0$ then

$$4a > \omega_1(2a - \mu). \tag{38}$$

Next, we distinguish three subcases: (a) $a \geq \frac{1}{2}\mu_{max}$ (b) $a \leq \frac{1}{2}\mu_{min}$ (c) $\frac{1}{2}\mu_{min} < a < \frac{1}{2}\mu_{max}$.
(a) $a \geq \frac{1}{2}\mu_{max}$. From (38) we have

$$\omega_1 < \frac{4a}{2a - \mu_{min}}. \tag{39}$$

Combining (31) and (39), it follows that

$$0 < \omega_1 < min\left\{2, \frac{4a}{2a - \mu_{min}}\right\},$$

or

$$0 < \omega_1 < 2. \tag{40}$$

Moreover, from (37) we have

$$0 < \omega_2 < \frac{2(2 - \omega_1)}{\omega_1 \mu_{max} + 2a(2 - \omega_1)} \tag{41}$$

which, because of (36), becomes

$$0 < \omega_2 < min\left\{\frac{1}{a}, \frac{2(2 - \omega_1)}{\omega_1 \mu_{max} + 2a(2 - \omega_1)}\right\}, \tag{42}$$

which yields (41) again. Therefore, for case (a) we have that (40) and (41) hold.
(b) $a \leq \frac{1}{2}\mu_{min}$. From (38) we have

$$\omega_1 > \frac{4a}{2a - \mu_{max}}. \tag{43}$$

Combining (31) and (43), it follows that

$$max\left\{0, \frac{4a}{2a - \mu_{max}}\right\} < \omega_1 < 2,$$

which yields (40). Therefore, for case (b) we have that (40) and (41) hold also as in case (a).

(c) $\frac{1}{2}\mu_{min} < a < \frac{1}{2}\mu_{max}$. Let α, β be two positive integers such that $\mu_\alpha = \max\{\mu|\mu \leq 2\alpha\}$, $\mu_\beta = \min\{\mu|\mu \geq 2\alpha\}$. Next, we distinguish two cases: (i) $\mu_{min} \leq \mu \leq \mu_\alpha$, (ii) $\mu_\beta \leq \mu \leq \mu_{max}$.

Case (i): $\mu_{min} \leq \mu \leq \mu_\alpha$. Following a similar approach as in Case (a), we have that (40) holds and

$$0 < \omega_2 < \frac{2(2 - \omega_1)}{\omega_1\mu_\alpha + 2a(2 - \omega_1)}. \tag{44}$$

Case (ii): $\mu_\beta \leq \mu \leq \mu_{max}$. Following a similar approach as in Case (b), we have that (40) holds and

$$0 < \omega_2 < \frac{2(2 - \omega_1)}{\omega_1\mu_{max} + 2a(2 - \omega_1)}. \tag{45}$$

Combining (44) and (45) it follows that

$$0 < \omega_2 < \min\left\{\frac{2(2 - \omega_1)}{\omega_1\mu_\alpha + 2a(2 - \omega_1)}, \frac{2(2 - \omega_1)}{\omega_1\mu_{max} + 2a(2 - \omega_1)}\right\}, \tag{46}$$

which is equivalent to (41). Hence, case 1 of table 1 is proved. Following a similar treatment we can prove the rest of the cases of Table 1. □

Corollary 2.3 *Under the hypothesis of Theorem 2.2 and if $a = 0$ then $\rho(\mathcal{L}(\omega_1, \omega_2)) < 1$ if*

$$0 < \omega_1 < 2 \quad and \quad 0 < \omega_2 < \frac{2(2 - \omega_1)}{\omega_1\mu_{max}}. \tag{47}$$

Proof If we let $a = 0$ in (21) and follow a similar approach as in the proof of Theorem 2.2 we can verify that (47) holds. □

Note that (47) was also obtained in [8]. The following corollary gives sufficient conditions for the GBSOR(a) method to converge.

Corollary 2.4 *Under the hypothesis of Theorem 2.2, $\rho(\mathcal{M}(\omega_1, \omega_2, a)) < 1$ if the parameters ω_1 and ω_2 lie in any case of Table 2.*

Proof Using the functional relationship (30) and following a similar approach as in the proof of Theorem 2.2 we have

$$0 < \omega_1 < 2 \text{ and } 0 < \frac{\omega_2}{1 - (1 - a)\omega_2} < \frac{2(2 - \omega_1)}{\omega_1\mu}. \tag{48}$$

Note that the second part of (48) is the same as (35) where now $1 - a$ appears instead of a. This occurs because the preconditioning matrix R is given by (22) and $\mathcal{U}$ is expressed in (4) in terms of $1 - a$. Therefore, if we let $1 - a$ in place of a in Table 1, we obtain Table 2. □

Table 2 Sufficient conditions for the GBSOR(a) method to converge if $\mu_{min} > 0$.

Condition	Cases	$\omega_2 - $ Domain	$\omega_1 - $ Domain
$a < 1$	1	$0 < \omega_2 < \dfrac{2(2 - \omega_1)}{\omega_1 \mu_{max} + 2(1 - a)(2 - \omega_1)}$	$0 < \omega_1 < 2$
$1 < a$	2	$\omega_2 < \dfrac{2(2 - \omega_1)}{\omega_1 \mu_{max} + 2(1 - a)(2 - \omega_1)} < 0$	$0 < \omega_1 < \dfrac{4(1 - a)}{2(1 - a) - \mu_{max}}$
	3	$0 < \omega_2$	$0 < \omega_1 < \dfrac{4(1 - a)}{2(1 - a) - \mu_{max}}$
	4	$0 < \omega_2 < \dfrac{2(2 - \omega_1)}{\omega_1 \mu_{max} + 2(1 - a)(2 - \omega_1)}$	$\dfrac{4(1 - a)}{2(1 - a) - \mu_{min}} < \omega_1 < 2$

Corollary 2.5 *Under the hypothesis of corollary 2.4 and if $a = 1$ then $\rho(\mathcal{M}(\omega_1, \omega_2, 1)) < 1$ if*

$$0 < \omega_1 < 2 \quad and \quad 0 < \omega_2 < \frac{2(2 - \omega_1)}{\omega_1 \mu_{max}}. \tag{49}$$

Proof If we let $a = 1$ in (48) then (49) follows immediately. $\qquad\square$

If the matrix Q is symmetric negative definite, then we have the following theorem.

Theorem 2.3 *Let $A \in \mathbb{R}^{m \times m}$ be symmetric positive definite, $B \in \mathbb{R}^{m \times n}$ be of full column rank and $Q \in \mathbb{R}^{n \times n}$ be symmetric negative definite. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1} B^T A^{-1} B$ by μ_{min} and μ_{max}, respectively. Then $\rho(\mathcal{L}(\omega_1, \omega_2, a)) < 1$ if the parameters ω_1 and ω_2 lie in the following cases of Table 3.*

Table 3 Sufficient conditions for the GSOR(a) method to converge if $\mu_{max} < 0$.

Condition	Cases	$\omega_2 - $ Domain	$\omega_1 - $ Domain
$a < 0$	1	$\dfrac{2(2 - \omega_1)}{\omega_1 \mu_{min} + 2a(2 - \omega_1)} < \omega_2 < 0$	$0 < \omega_1 < 2$
$a > 0$	2	$(0 <) \dfrac{2(2 - \omega_1)}{\omega_1 \mu_{min} + 2a(2 - \omega_1)} < \omega_2$	$0 < \omega_1 < \dfrac{4a}{2a - \mu_{min}}$
	3	$\omega_2 < 0$	$0 < \omega_1 < \dfrac{4a}{2a - \mu_{min}}$
	4	$\dfrac{2(2 - \omega_1)}{\omega_1 \mu_{min} + 2a(2 - \omega_1)} < \omega_2 < 0$	$\dfrac{4a}{2a - \mu_{max}} < \omega_1 < 2$

Proof Using the functional relationship (21) and following a similar approach as in the proof of Theorem 2.2 taking into consideration that $\mu_{max} < 0$ we have

$$0 < \omega_1 < 2 \quad and \quad \frac{2(2 - \omega_1)}{\omega_1 \mu} < \frac{\omega_2}{1 - a\omega_2} < 0. \tag{50}$$

From (50) the cases presented in Table 3 can be readily verified. $\qquad\square$

Corollary 2.6 *Under the hypothesis of Theorem 2.3 and if $a = 0$ then $\rho(\mathcal{L}(\omega_1, \omega_2)) < 1$ if*

$$0 < \omega_1 < 2 \quad and \quad \frac{2(2 - \omega_1)}{\omega_1 \mu_{min}} < \omega_2 < 0. \tag{51}$$

Proof Using the functional relationship (21) and following the proof of Theorem 2.3 we have that if $a = 0$ in (50) then (51) follows.   □

The above result was also obtained in [8].

Theorem 2.4 *Let $A \in \mathbb{R}^{m \times m}$ be symmetric positive definite, $B \in \mathbb{R}^{m \times n}$ be of full column rank and $Q \in \mathbb{R}^{n \times n}$ be symmetric negative definite. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1}B$ by μ_{min} and μ_{max}, respectively. Then $\rho(\mathcal{M}(\omega_1, \omega_2, a)) < 1$ if the parameters ω_1 and ω_2 lie in the following cases of Table 3.*

Table 4 Sufficient conditions for the GBSOR(a) method to converge if $\mu_{max} < 0$.

Condition	Cases	$\omega_2 - \text{Domain}$	$\omega_1 - \text{Domain}$
$a < 1$	1	$\dfrac{2(2-\omega_1)}{\omega_1 \mu_{min} + 2(1-a)(2-\omega_1)} < \omega_2 < 0$	$0 < \omega_1 < 2$
$a > 1$	2	$(0 <)\dfrac{2(2-\omega_1)}{\omega_1 \mu_{min} + 2(1-a)(2-\omega_1)} < \omega_2$	$0 < \omega_1 < \dfrac{4(1-a)}{2(1-a) - \mu_{min}}$
	3	$\omega_2 < 0$	$0 < \omega_1 < \dfrac{4(1-a)}{2(1-a) - \mu_{min}}$
	4	$\dfrac{2(2-\omega_1)}{\omega_1 \mu_{min} + 2(1-a)(2-\omega_1)} < \omega_2 < 0$	$\dfrac{4(1-a)}{2(1-a) - \mu_{max}} < \omega_1 < 2$

Proof Using the functional relationship (30) and following a similar approach as in the proof of Theorem 2.2 taking into consideration that $\mu_{max} < 0$ we have

$$0 < \omega_1 < 2 \quad \text{and} \quad \frac{2(2-\omega_1)}{\omega_1 \mu} < \frac{\omega_2}{1 - (1-a)\omega_2} < 0. \tag{52}$$

From (52) the cases presented in Table 4 can be readily verified.   □

Corollary 2.7 *Under the hypothesis of Theorem 2.4 and if $a = 1$ then $\rho(\mathcal{M}(\omega_1, \omega_2)) < 1$ if*

$$0 < \omega_1 < 2 \quad \text{and} \quad \frac{2(2-\omega_1)}{\omega_1 \mu_{min}} < \omega_2 < 0. \tag{53}$$

Proof Using the functional relationship (30) and following the proof of Theorem 2.4 we have that if $a = 1$ in (52) then (53) follows.   □

In the sequel we study the convergence analysis of the GMESOR method under the same assumptions.

2.2.2 The GMESOR method

The next theorem provides sufficient conditions for the GMESOR method to converge if the matrix Q is symmetric positive definite and $a = 0$. The study of the case $a \neq 0$ follows a similar but cumbersome approach as it requires many cases to be examined. This study will not have any substantial contribution since the minimum value of the spectral radius of the GMESOR(a) method is independent of a (Theorem 2.10), meaning that for, say $a = 0$, the GMESOR

method will attain the maximum rate of convergence. So, we are interested to find the convergence ranges of the parameters of the GMESOR(a) method for the simplified case when $a = 0$.

Theorem 2.5 *Consider the GMESOR method. Let $A \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m \times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1} B^T A^{-1} B$ by μ_{min} and μ_{max}, respectively. Then $\rho(\mathcal{H}(\tau_1, \tau_2, \omega_2)) < 1$ if*

$$0 < \tau_1 < 2, \quad 0 < \tau_2 < \bar{\tau}_2(\mu_{max}) \quad and \quad \underline{\omega}_2(\mu_{max}) < \omega_2 < \bar{\omega}_2(\mu_{max}) \qquad (54)$$

where

$$\bar{\tau}_2(\mu_{max}) = \frac{4}{\tau_1 \mu_{max}}, \quad \underline{\omega}_2(\mu_{max}) = \tau_2 - \frac{1}{\mu_{max}} \quad and \quad \bar{\omega}_2(\mu_{max}) = \frac{2 - \tau_1}{\tau_1 \mu_{max}} + \frac{\tau_2}{2}. \qquad (55)$$

Proof Recall that $\lambda = 1 - \tau_1 \neq 0$ is an eigenvalue of $\mathcal{H}(\tau_1, \tau_2, \omega_2)$ and if $\lambda \neq 1 - \tau_1$ then the eigenvalues of $\mathcal{H}(\tau_1, \tau_2, \omega_2)$ are given by (13) where $a = 0$. If $\lambda = 1 - \tau_1 \neq 0$, then the GMESOR method is convergent if and only if $|\lambda| < 1$, that is $|1 - \tau_1| < 1$, or

$$0 < \tau_1 < 2, \qquad (56)$$

which is the first inequality of (54). If $\lambda \neq 1 - \tau_1$, then (13) holds and by Lemma 2.1 page 171 of [55], it follows that the GMESOR method is convergent if and only if (136) holds where

$$c = 1 - \tau_1 + \tau_1(\tau_2 - \omega_2)\mu \qquad (57)$$

and

$$b = 2 - \tau_1 - \tau_1 \omega_2 \mu. \qquad (58)$$

From the first inequality of (136) it follows that

$$0 < 1 + c < 2. \qquad (59)$$

From the second inequality of (136), because of (57) and (58), we have

$$|1 + c - \tau_1 \tau_2 \mu| < 1 + c$$

or

$$0 < \frac{\tau_1 \tau_2 \mu}{2} < 1 + c. \qquad (60)$$

Combining (139) and (60), it follows that

$$0 < \frac{\tau_1 \tau_2 \mu}{2} < 1 + c < 2. \qquad (61)$$

In order for (61) to hold we must have

$$0 < \frac{\tau_1 \tau_2 \mu}{2} < 2$$

or, because of (56),

$$0 < \tau_2 < \frac{4}{\tau_1 \mu}, \tag{62}$$

which proves the second inequality of (54). Inequality (61), because of (57), becomes

$$\frac{\tau_1 \tau_2 \mu}{2} < 2 - \tau_1 + \tau_1 \tau_2 \mu - \tau_1 \omega_2 \mu < 2$$

which is equivalent to

$$\tau_2 - \frac{1}{\mu} < \omega_2 < \frac{2 - \tau_1}{\tau_1 \mu} + \frac{\tau_2}{2}. \tag{63}$$

By studying the monotonicity of the right and left hand side of (63) with respect to μ we obtain the third inequality of (54). $\qquad \square$

The convergence conditions for GESOR are given by the following corollary.

Corollary 2.8 *Consider the GESOR method. Let $A \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m \times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1} B^T A^{-1} B$ by μ_{min} and μ_{max}, respectively. Then $\rho(\mathcal{H}(\tau, \omega_2)) < 1$ if*

$$0 < \tau < \bar{\tau}(\mu_{max}) \quad and \quad \underline{\omega}_2(\tau) < \omega_2 < \bar{\omega}_2(\tau), \tag{64}$$

where

$$\bar{\tau}(\mu_{max}) = \begin{cases} 2, & \mu_{max} \leq 1 \\ \frac{2}{\sqrt{\mu_{max}}}, & \mu_{max} > 1, \end{cases} \tag{65}$$

$$\underline{\omega}_2(\tau) = \tau - \frac{1}{\mu_{max}} \quad and \quad \bar{\omega}_2(\tau) = \frac{2 - \tau}{\tau \mu_{max}} + \frac{\tau}{2}. \tag{66}$$

Proof Letting $\tau = \tau_1 = \tau_2$ in (54) we obtain (64). $\qquad \square$

The convergence area for the GESOR method is illustrated in figure 1. Note that as μ_{max} increases the point of intersection of the two curves $\bar{\omega}_2(\tau)$ and $\underline{\omega}_2(\tau)$ moves towards zero and the convergence area of the GESOR method shrinks. However, in practice μ_{max} usually is < 1.

If the matrix Q is symmetric negative definite and $a = 0$ then we have the following theorem.

Theorem 2.6 *Consider the GMESOR method. Let $A \in \mathbb{R}^{m \times m}$ be symmetric positive definite, $B \in \mathbb{R}^{m \times n}$ be of full column rank and $Q \in \mathbb{R}^{n \times n}$ be symmetric negative definite. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1} B^T A^{-1} B$ by μ_{min} and μ_{max}, respectively. Then $\rho(\mathcal{H}(\tau_1, \tau_2, \omega_2)) < 1$ if*

$$0 < \tau_1 < 2, \quad \underline{\tau}_2(\mu_{min}) < \tau_2 < 0 \quad and \quad \underline{\omega}_2(\mu_{min}) < \omega_2 < \bar{\omega}_2(\mu_{min}). \tag{67}$$

where

$$\underline{\tau}_2(\mu_{min}) = \frac{4}{\tau_1 \mu_{min}}, \quad \underline{\omega}_2(\mu_{min}) = \frac{2 - \tau_1}{\tau_1 \mu_{min}} + \frac{\tau_2}{2} \quad and \quad \bar{\omega}_2(\mu_{min}) = \tau_2 - \frac{1}{\mu_{min}}. \tag{68}$$
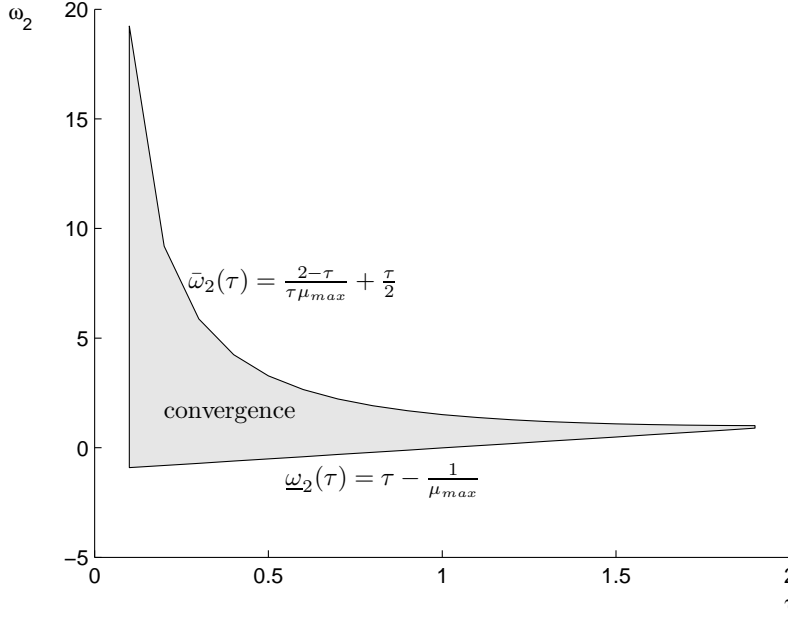
Fig. 1 Convergence area of the GESOR method for $\mu_{max} = 0.99$.

Proof Following a similar approach as in the proof of Theorem 2.5 and using the functional relationship (13) we can prove (67). □

Corollary 2.9 *Consider the GESOR method. Let $A \in \mathbb{R}^{m \times m}$ be symmetric positive definite, $B \in \mathbb{R}^{m \times n}$ be of full column rank and $Q \in \mathbb{R}^{n \times n}$ be symmetric negative definite. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1}B$ by μ_{min} and μ_{max}, respectively. Then $\rho(\mathcal{H}(\tau, \omega_2)) < 1$ if*

$$\underline{\tau}_1(\mu_{min}) < \tau < 2, \quad \underline{\omega}_2(\mu_{min}) < \omega_2 < \bar{\omega}_2(\mu_{min}) \quad and \quad \mu_{min} > 1, \qquad (69)$$

where

$$\underline{\tau}_1(\mu_{min}) = \frac{2}{\sqrt{\mu_{min}}}, \quad \underline{\omega}_2(\mu_{min}) = \frac{2-\tau}{\tau\mu_{min}} + \frac{\tau}{2} \quad and \quad \bar{\omega}_2(\mu_{min}) = \tau - \frac{1}{\mu_{min}}. \qquad (70)$$

Proof (69) is proved by following a similar approach as in the proof of Theorem 2.5 and using the functional relationship (19). □

2.3 Optimum parameters

In this section we determine optimum values for the parameters of the iterative methods studied in the present section under the hypothesis that $a \neq 0$ and

the eigenvalues of the matrix J are real. We assume that Q is a symmetric positive or negative definite matrix.

2.3.1 The GSOR(a) method

In the following theorem the optimum parameters for the GSOR(a) method are determined assuming that the matrix Q is symmetric positive definite.

Theorem 2.7 *Consider the GSOR(a) method. Let $A \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m \times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1} B$ by μ_{min} and μ_{max}, respectively. Then the spectral radius of the GSOR(a) method, $\rho(\mathcal{L}(\omega_1, \omega_2, a))$, is minimized for any $a \neq -\sqrt{\mu_{min}\mu_{max}}$ at*

$$\omega_{1_{opt}} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2} \quad and \quad \omega_{2_{opt}} = \frac{1}{a + \sqrt{\mu_{min}\mu_{max}}} \tag{71}$$

and its corresponding value is

$$\rho(\mathcal{L}(\omega_{1_{opt}}, \omega_{2_{opt}}, a)) = \frac{\sqrt{\mu_{max}} - \sqrt{\mu_{min}}}{\sqrt{\mu_{max}} + \sqrt{\mu_{min}}}. \tag{72}$$

Proof The functional relationship (21) may be written as follows

$$(\lambda + \omega_1 - 1)(\lambda - 1) = -\lambda\omega_1\hat{\omega}_2\mu \tag{73}$$

where

$$\hat{\omega}_2 = \frac{\omega_2}{1 - a\omega_2} \tag{74}$$

with $a\omega_2 \neq 1$. The optimum values of ω_1 and $\hat{\omega}_2$ will be determined such that

$$\rho(\mathcal{L}(\omega_1, \omega_2, a)) = \max_{\mu_{min} \leq \mu \leq \mu_{max}} |\lambda| \tag{75}$$

is minimum. The real roots of (73) are the intersection points of the parabola

$$g_{\omega_1}(\lambda) = \frac{(\lambda + \omega_1 - 1)(\lambda - 1)}{\omega_1\hat{\omega}_2} \tag{76}$$

and the straight lines

$$h(\lambda) = -\lambda\mu, \;\; 0 < \mu_{min} \leq \mu \leq \mu_{max}. \tag{77}$$

Following a similar argument as in [50] page 111, $h(\lambda)$ are straight lines through the point $(0,0)$ and $g_{\omega_1}(\lambda)$ is a parabola passing through the point $(1,0)$. The discriminant of (21) is

$$\Delta(\omega_1, \hat{\omega}_2, \mu) = (2 - \omega_1 - \omega_1\hat{\omega}_2\mu)^2 - 4(1 - \omega_1). \tag{78}$$

Note that $\Delta(\omega_1, \hat{\omega}_2, \mu) \leq 0$ for $0 < \omega_1 \leq \tilde{\omega}_1(\mu)$ and $\Delta(\omega_1, \hat{\omega}_2, \mu) \geq 0$ for $\tilde{\omega}_1(\mu) \leq \omega_1 < 2$, where

$$\tilde{\omega}_1(\mu) = \frac{4\hat{\omega}_2\mu}{(1 + \hat{\omega}_2\mu)^2}. \tag{79}$$

If $0 < \omega_1 \le \tilde{\omega}_1(\mu)$ then the value of $\rho(\mathcal{L}(\omega_1, \omega_2, a))$ is

$$|\tilde{\lambda}_1| = |\tilde{\lambda}_N| = (1 - \omega_1)^{1/2}, \tag{80}$$

where $\tilde{\lambda}_1$ and $\tilde{\lambda}_N$ are the two conjugate complex roots of (21). Furthermore, (80) is a decreasing function of ω_1. In case $\tilde{\omega}_1(\mu) \le \omega_1 < 2$ the roots of (21) can be geometrically interpreted as the intersection of the curves $g_{\omega_1}(\lambda)$ and $h(\lambda) = -\lambda\mu$, as illustrated in figure 2, where we have assumed, without loss of generality, that $h(\lambda) \equiv h_1(\lambda) = -\lambda\mu_{max}$. The largest abscissa of the two
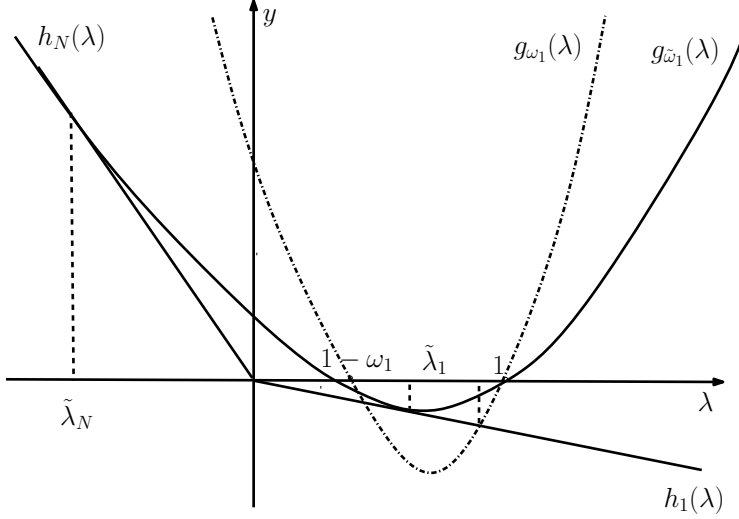

Fig. 2 Graphs of $g_{\omega_1}(\lambda), h_1(\lambda)$ and $h_N(\lambda)$ in case the roots of (73) are real.

points of intersection of $h(\lambda)$ and $g_{\omega_1}(\lambda)$ decreases with increasing ω_1. Indeed as ω_1 increases, the intersection point $(1 - \omega_1, 0)$ of $g_{\omega_1}(\lambda)$ with the $O\lambda$ axis is moving towards to zero until $g_{\omega_1}(\lambda)$ becomes tangent to $h(\lambda)$. Thus, for the fixed eigenvalue μ of J, the value of ω_1 which minimizes the zero of largest modulus of (21) is $\tilde{\omega}_1(\mu)$. Note that the straight lines $h_1(\lambda) = -\lambda\mu_{max}$ and $h_N(\lambda) = -\lambda\mu_{min}$ include all the lines $h(\lambda) = -\lambda\mu$. Therefore, (79) yields the two optima $\tilde{\omega}_1(\mu_{max})$ and $\tilde{\omega}_1(\mu_{min})$. However, these values must be equal as there is only one optimum, hence

$$\frac{4\hat{\omega}_2\mu_{max}}{(1 + \hat{\omega}_2\mu_{max})^2} = \frac{4\hat{\omega}_2\mu_{min}}{(1 + \hat{\omega}_2\mu_{min})^2} \tag{81}$$

or

$$\hat{\omega}_2 = \frac{1}{\sqrt{\mu_{min}\mu_{max}}} \tag{82}$$

which, because of (74), yields the optimum value for ω_2 given by the second part of (71). Substituting the value of $\hat{\omega}_2$ in the expressions $\tilde{\omega}_1(\mu_{max})$ or

$\tilde{\omega}_1(\mu_{min})$, given by the first or second part of the equality (81), respectively, we obtain the optimum value of ω_1 given by the first part of (71). The spectral radius is given by

$$\rho(\mathcal{L}(\omega_1, \omega_2, a)) = \max\{|\tilde{\lambda}_1|, |\tilde{\lambda}_N|\} \tag{83}$$

where $\tilde{\lambda}_1, \tilde{\lambda}_N$ are the abscissas of the points of tangent of $h_1(\lambda), h_N(\lambda)$, respectively. For the minimization of $\rho(\mathcal{L}(\omega_1, \omega_2, a))$ with respect to ω_1 we require

$$|\tilde{\lambda}_1| = |\tilde{\lambda}_N|$$

or

$$\tilde{\lambda}_1 = -\tilde{\lambda}_N = (1 - \tilde{\omega}_1(\mu_{max}))^{1/2}, \tag{84}$$

where the last equality holds by the fact that $\tilde{\lambda}_1$, $\tilde{\lambda}_N$ are the abscissas of the tangents $h_1(\lambda)$ and $h_N(\lambda)$, respectively. From (80) and (84) it follows that

$$\rho(\mathcal{L}(\omega_1, \omega_2, a)) = (1 - \omega_{1_{opt}})^{1/2}$$

which, because of (71), yields (72). $\square$

Theorem 2.7 finds the optimum values of the relaxation parameters ω_1 and ω_2 of the GSOR(a) method. Letting $a = 0$ in (71) we obtain the optima found also in [8]. Note that the parameter a has no impact on the spectral radius of the GSOR(a) method as one might have expected. The algebraic approach in [8] is similar to the one followed by [55] for determining the optimum of the sole parameter in the SOR method. In case of GSOR(a), which has two parameters, there is an alternative less tedious algebraic approach (see [55] pp. 279-281). However, it remains to be verified whether either approach can be used to solve the problem of determining the optimum values of more than two parameters as is the case for the GMESOR(a) method. Our approach follows the geometric approach of Varga [50] for the determination of the optimum value of the parameter ω in SOR. It should be noted that this approach is also mentioned in [47] but without a proof.

Corollary 2.10 *Consider the GBSOR(a) method. Under the hypothesis of Theorem 2.7 the spectral radius of the GBSOR(a) method, $\rho(\mathcal{M}(\omega_1, \omega_2, a))$, is minimized for any $a \neq 1 + \sqrt{\mu_{min}\mu_{max}}$ at*

$$\omega_{1_{opt}} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2} \quad and \quad \omega_{2_{opt}} = \frac{1}{(1-a) + \sqrt{\mu_{min}\mu_{max}}} \tag{85}$$

and its corresponding value is

$$\rho(\mathcal{M}(\omega_{1_{opt}}, \omega_{2_{opt}}, a)) = \frac{\sqrt{\mu_{max}} - \sqrt{\mu_{min}}}{\sqrt{\mu_{max}} + \sqrt{\mu_{min}}}. \tag{86}$$

Proof We remark that the functional relationship (30) of GBSOR(a) is the same as that of the GSOR(a) method (21) with the only difference that now we have $1 - a$ instead of a. Therefore, we have the same results as in Theorem 2.7, if we simply replace a with $1 - a$. $\qquad\square$

If the matrix Q is symmetric negative definite, the optimum parameters and the minimum spectral radius for the GSOR(a) method are given by the following theorem.

Theorem 2.8 *Consider the GSOR(a) method. Let $A \in \mathbb{R}^{m \times m}$ be symmetric positive definite, $B \in \mathbb{R}^{m \times n}$ be of full column rank and $Q \in \mathbb{R}^{n \times n}$ be symmetric negative definite. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1}B$ by μ_{min} and μ_{max}, respectively. Then the spectral radius of the GSOR(a) method $\rho(\mathcal{L}(\omega_1, \omega_2, a))$, when the matrix J has negative eigenvalues, is minimized for any $a \neq \sqrt{\mu_{min}\mu_{max}}$ at*

$$\omega_{1_{opt}} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{|\mu_{min}|} + \sqrt{|\mu_{max}|})^2} \quad and \quad \omega_{2_{opt}} = \frac{1}{a - \sqrt{\mu_{min}\mu_{max}}} \tag{87}$$

and its corresponding value is

$$\rho(\mathcal{L}(\omega_{1_{opt}}, \omega_{2_{opt}}, a)) = \frac{\sqrt{|\mu_{min}|} - \sqrt{|\mu_{max}|}}{\sqrt{|\mu_{min}|} + \sqrt{|\mu_{max}|}}. \tag{88}$$

Proof In this case $\mu < 0$. Following a similar approach as in Theorem 2.7, we obtain (87) and (88). $\qquad\square$

Under the hypothesis of Theorem 2.8 and if $a = 0$, these results were also obtained in [8].

Theorem 2.9 *Consider the GBSOR(a) method. Let $A \in \mathbb{R}^{m \times m}$ be symmetric positive definite, $B \in \mathbb{R}^{m \times n}$ be of full column rank and $Q \in \mathbb{R}^{n \times n}$ be symmetric negative definite. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1}B$ by μ_{min} and μ_{max}, respectively. Then the spectral radius of the GBSOR(a) method $\rho(\mathcal{M}(\omega_1, \omega_2, a))$, when the matrix J has negative eigenvalues, is minimized for any $a \neq 1 - \sqrt{\mu_{min}\mu_{max}}$ at*

$$\omega_{1_{opt}} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{|\mu_{min}|} + \sqrt{|\mu_{max}|})^2} \quad and \quad \omega_{2_{opt}} = \frac{1}{1 - a - \sqrt{\mu_{min}\mu_{max}}} \tag{89}$$

and its corresponding value is

$$\rho(\mathcal{L}(\omega_{1_{opt}}, \omega_{2_{opt}}, a)) = \frac{\sqrt{|\mu_{min}|} - \sqrt{|\mu_{max}|}}{\sqrt{|\mu_{min}|} + \sqrt{|\mu_{max}|}}. \tag{90}$$

Proof We remark that the functional relationship (30) of GBSOR(a) is the same as that of the GSOR(a) method (21) with the only difference that now we have $1 - a$ instead of a. Therefore, we can apply the results of Theorem 2.8 by replacing a with $1 - a$. $\qquad\square$

2.3.2 The GMESOR(a) method

In the sequel we determine the optimum parameters for the GMESOR(a) method.

Theorem 2.10 *Consider the GMESOR(a) method. Let $A \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m \times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1}B$ by μ_{min} and μ_{max}, respectively. Then the spectral radius of the GMESOR(a) method, $\rho(\mathcal{H}(\tau_1, \tau_2, \omega_2, a))$, is minimized for any $a \neq -\sqrt{\mu_{min}\mu_{max}}$ at*

$$\omega_{2_{opt}} = \tau_{2_{opt}}, \tag{91}$$

$$\tau_{1_{opt}} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2} \quad and \quad \tau_{2_{opt}} = \frac{1}{a + \sqrt{\mu_{min}\mu_{max}}} \tag{92}$$

and its corresponding value is

$$\rho(\mathcal{H}(\tau_{1_{opt}}, \tau_{2_{opt}}, \omega_{2_{opt}}, a)) = \frac{\sqrt{\mu_{max}} - \sqrt{\mu_{min}}}{\sqrt{\mu_{max}} + \sqrt{\mu_{min}}}. \tag{93}$$

Proof The functional relationship of the GMESOR(a) method is given by (13) or

$$(\lambda + \tau_1 - 1)(\lambda - 1) = \frac{\tau_1(\omega_2 - \tau_2 - \lambda\omega_2)\mu}{1 - a\omega_2}. \tag{94}$$

The optimum values of τ_1, τ_2 and ω_2 will be determined such that

$$\rho(\mathcal{H}(\tau_1, \tau_2, \omega_2, a)) = \max_{\mu_{min} \leq \mu \leq \mu_{max}} |\lambda| \tag{95}$$

is minimum. The real roots of (13) are the intersection points of the parabola

$$g(\lambda) = \frac{(\lambda + \tau_1 - 1)(\lambda - 1)}{\tau_1} \tag{96}$$

and the straight lines

$$h(\lambda) = \frac{\omega_2 - \tau_2 - \lambda\omega_2}{1 - a\omega_2}\mu, \quad 0 < \mu_{min} \leq \mu \leq \mu_{max}. \tag{97}$$

Following a similar argument as in [50] page 111, $h(\lambda)$ are straight lines through the point $\left(0, \frac{\omega_2 - \tau_2}{1 - a\omega_2}\mu\right)$ and $g_{\tau_1}(\lambda)$ is a parabola passing through the points (1,0) and $(1 - \tau_1, 0)$ (see figure 3). The spectral radius is given by

$$\rho(\mathcal{H}(\tau_1, \tau_2, \omega_2, a)) = \max\{|\tilde{\lambda}_1|, |\tilde{\lambda}_N|\} \tag{98}$$

where $\tilde{\lambda}_1, \tilde{\lambda}_N$ are the abscissas of the points of tangent of $h_1(\lambda), h_N(\lambda)$, respectively, where now $h_1(\lambda) = (\omega_2 - \tau_2 - \lambda\omega_2)\mu_{max}$ and $h_N(\lambda) = (\omega_2 - \tau_2 - \lambda\omega_2)\mu_{min}$. Therefore,

$$|\tilde{\lambda}_1| = \left(1 - \tau_1 + \tau_1\frac{\tau_2 - \omega_2}{1 - a\omega_2}\mu_{max}\right)^{1/2} \tag{99}$$
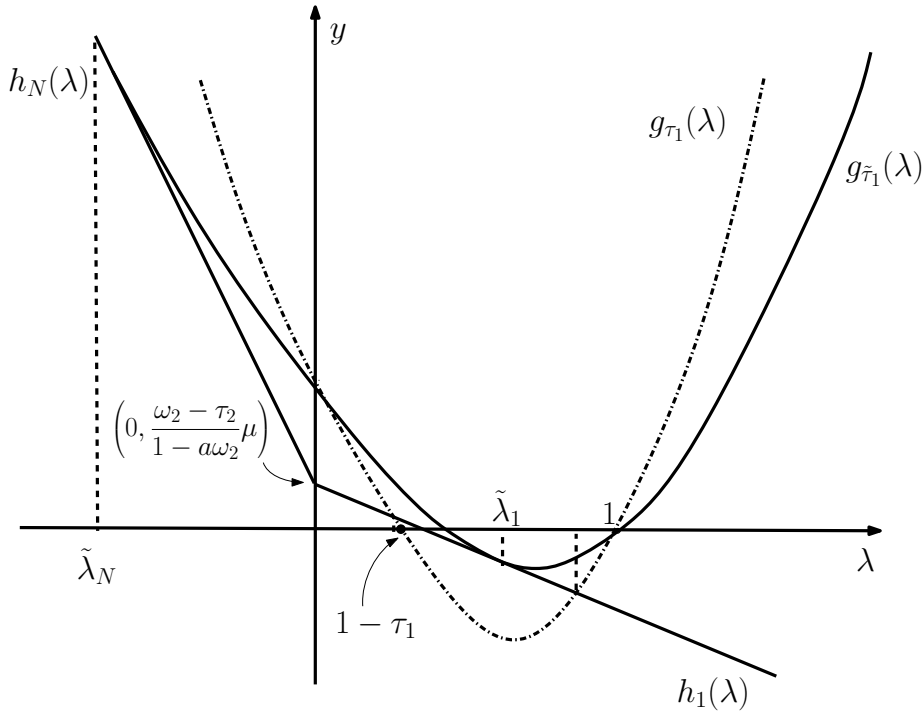
Fig. 3 Graphs of $g_{\tau_1}(\lambda), h_1(\lambda)$ and $h_N(\lambda)$ in case the roots of (94) are real.

and

$$|\tilde{\lambda}_N| = \left(1 - \tau_1 + \tau_1 \frac{\tau_2 - \omega_2}{1 - a\omega_2}\mu_{min}\right)^{1/2}. \tag{100}$$

From (98) it follows that the minimum value of $\rho(\mathcal{H}(\tau_1, \tau_2, \omega_2, a))$ is attained when

$$|\tilde{\lambda}_1| = |\tilde{\lambda}_N| \tag{101}$$

which, because of (99) and (100), implies

$$\omega_2 = \tau_2. \tag{102}$$

In case $\tilde{\lambda}_1$ and $\tilde{\lambda}_N$ are the two conjugate complex roots of (94), it follows that (101) must also hold for $\rho(\mathcal{H}(\tau_1, \tau_2, \omega_2, a))$ to be minimized. So, (102) holds if either (94) has real or conjugate complex roots. However, if (102) holds, then (13) becomes

$$\lambda^2 + \lambda\left(\tau_1 - 2 + \tau_1\hat{\tau}_2\mu\right) + 1 - \tau_1 = 0,$$

which is the functional relationship of the GSOR with

$$\hat{\tau}_2 = \frac{\tau_2}{1 - a\tau_2}. \tag{103}$$

Therefore the optimum values of τ_1 and $\hat{\tau}_2$ are given by ω_{opt} and τ_{opt} of [8], respectively, whereas the minimum value of $\rho(\mathcal{H}(\tau_1, \tau_2, \omega_2, a))$ is given by $\rho(\mathcal{H}(\omega_{opt}, \tau_{opt}))$ of [8]. Finally, using (103) we find (92). $\qquad\square$
So, for the optimum values of its parameters, GMESOR(a) degenerates to the GSOR(a) method.

Corollary 2.11 *Consider the GESOR(a) method. Let $A \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m \times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1} B$ by μ_{min} and μ_{max}, respectively. Then the spectral radius of the GESOR(a) method, $\rho(\mathcal{H}(\tau, \omega_2, a))$, is minimized at*

$$\omega_{2_{opt}} = \tau_{opt} \tag{104}$$

and

$$\tau_{opt} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2}, \quad a_{opt} = \frac{1}{\tau_{opt}} - \sqrt{\mu_{max}\mu_{min}} \tag{105}$$

and its corresponding value is

$$\rho(\mathcal{H}(\tau_{opt}, \omega_{2_{opt}}, a_{opt})) = \frac{\sqrt{\mu_{max}} - \sqrt{\mu_{min}}}{\sqrt{\mu_{max}} + \sqrt{\mu_{min}}}. \tag{106}$$

Proof Recall that GESOR(a) is obtained by setting $\tau_1 = \tau_2$ in GMESOR(a). Therefore, (104) and (105) are obtained by (91) and (92), respectively, where now we require $\tau_{opt} = \tau_{1_{opt}} = \tau_{2_{opt}}$. $\qquad\square$

Corollary 2.12 *Consider the GMEBSOR(a) method. Under the hypothesis of Theorem 2.10 the spectral radius of the GMEBSOR(a) method, $\rho(\mathcal{K}(\tau_1, \tau_2, \omega_1, \omega_2, a))$, is minimized at*

$$\omega_{1_{opt}} = \tau_{1_{opt}}, \tag{107}$$

where

$$\tau_{1_{opt}} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2} \quad and \quad \tau_{2_{opt}} = \frac{1 - (1-a)\omega_2}{\sqrt{\mu_{min}\mu_{max}}} \tag{108}$$

and its corresponding value is

$$\rho(\mathcal{K}(\tau_{1_{opt}}, \tau_{2_{opt}}, \omega_{1_{opt}}, \omega_2, a)) = \frac{\sqrt{\mu_{max}} - \sqrt{\mu_{min}}}{\sqrt{\mu_{max}} + \sqrt{\mu_{min}}}. \tag{109}$$

Proof Following a similar approach as in Theorem 2.10, using the functional relationship (28) and requiring $|\tilde{\lambda}_1| = |\tilde{\lambda}_N|$ we find

$$\omega_1 = \tau_1. \tag{110}$$

Therefore, (28) because of (110) becomes

$$\lambda^2 + \lambda(\tau_1 - 2 + \tau_1\hat{\tau}_2\mu) + 1 - \tau_1 = 0 \tag{111}$$

with

$$\hat{\tau}_2 = \frac{\tau_2}{1 - (1-a)\omega_2}, \tag{112}$$

which is the functional relationship of the GSOR(a) method (see (21)) with the only difference that now we have $1 - a$ instead of a in $\hat{\tau}_2$, hence (108) and (109) hold because of Theorem 2.7. $\qquad\square$
Note that although the GMEBSOR(a) method has four parameters instead of three as in the GMESOR(a) method, both methods have the same minimum spectral radius.

Corollary 2.13 *Consider the GEBSOR(a) method. Under the hypothesis of Theorem 2.10 the spectral radius of the GEBSOR(a) method, $\rho(\mathcal{K}(\tau, \omega_2, a))$, is minimized at*

$$\omega_{1_{opt}} = \tau_{opt} \tag{113}$$

and

$$\tau_{opt} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2}, \quad \omega_{2_{opt}} = \frac{1 - \tau_{opt}\sqrt{\mu_{min}\mu_{max}}}{1 - a} \tag{114}$$

and its corresponding value is

$$\rho(\mathcal{K}(\tau_{opt}, \omega_{2_{opt}}, a)) = \frac{\sqrt{\mu_{max}} - \sqrt{\mu_{min}}}{\sqrt{\mu_{max}} + \sqrt{\mu_{min}}}. \tag{115}$$

Proof Recall that GEBSOR(a) is obtained by setting $\tau_1 = \tau_2$ in GMEBSOR(a). Therefore, (113) and (114) are obtained by (107) and (108), respectively, where now we require $\tau_{opt} = \tau_{1_{opt}} = \tau_{2_{opt}}$. $\qquad\square$

Our analysis so far shows that all the studied iterative methods (GMESOR(a), GMEBSOR(a)) have also the same rate of convergence as the PCG method for the optimum values of their parameters (see Theorems 2.7, 2.10 and corollary 2.12).

3 The Generalized Modified Preconditioned Simultaneous Displacement (GMPSD) method

The Preconditioned Simultaneous Displacement (PSD) method was introduced in [19]. When the coefficient matrix A is two-cyclic the Modified PSD (MPSD) method was studied in [34], [38]. Motivated by our previous work we introduce the Generalized Modified PSD (GMPSD) method and study its convergence rate for the numerical solution of the augmented linear system (1)-(2).

3.1 The functional relationship

In the sequel, we let the preconditioning matrix R be the product of the lower triangular part with the upper triangular part of $\mathcal{A}$ in an attempt to obtain a better approximation of $\mathcal{A}$ and consequently an increase in the rate of convergence of the corresponding iterative method. Let

$$R = (\mathcal{D} - \Omega\mathcal{L})\mathcal{D}^{-1}(\mathcal{D} - \Omega\mathcal{U}). \tag{116}$$

From (6) and (116) it follows that the iteration matrix of (5) now is

$$\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2, a) = I - (\mathcal{D} - \Omega\mathcal{U})^{-1} D (\mathcal{D} - \Omega\mathcal{L})^{-1} T \mathcal{A} \qquad (117)$$

whereas $\eta(\tau_1, \tau_2)$ in (6) corresponds to

$$\gamma(\tau_1, \tau_2, \omega_1, \omega_2, a) = (\mathcal{D} - \Omega\mathcal{U})^{-1} D (\mathcal{D} - \Omega\mathcal{L})^{-1} T b. \qquad (118)$$

Note that this method has four parameters τ_1, τ_2, ω_1 and ω_2 instead of three in the GMESOR method. The iterative scheme given by (5), (117) and (118) will be referred to as the Generalized Modified Preconditioned Simultaneous Displacement (GMPSD) method. For $(\mathcal{D} - \Omega\mathcal{U})^{-1} D (\mathcal{D} - \Omega\mathcal{L})^{-1}$ to exist we require

$$\det[(\mathcal{D} - \Omega\mathcal{L})\mathcal{D}^{-1}(\mathcal{D} - \Omega\mathcal{U})] \neq 0. \qquad (119)$$

Because of (4)

$$R = (\mathcal{D} - \Omega\mathcal{L})\mathcal{D}^{-1}(\mathcal{D} - \Omega\mathcal{U}) = \begin{pmatrix} A & \omega_1 B \\ -\omega_2 B^T & (1 - a\omega_2)[1 - (1-a)\omega_2]Q - \omega_1\omega_2 B^T A^{-1} B \end{pmatrix}.$$
$$(120)$$

Therefore,

$$\det(\mathcal{D} - \Omega\mathcal{L})\mathcal{D}^{-1}(\mathcal{D} - \Omega\mathcal{U}) = (1 - a\omega_2)^n [1 - (1-a)\omega_2]^n \det(A) \det(Q) \neq 0$$

or

$$a \neq \frac{1}{2} \text{ and } \omega_2 \neq 2 \qquad (121)$$

since the matrix A is symmetric positive definite and the matrix Q is nonsingular. The GMPSD method has the following algorithmic form.

THE GMPSD METHOD: *Let $Q \in \mathbb{R}^{n \times n}$ be a nonsingular and symmetric matrix. Given initial vectors $x^{(0)} \in \mathbb{R}^m$ and $y^{(0)} \in \mathbb{R}^n$, and relaxation factors $\tau_1, \tau_2 \neq 0$, $\omega_1, \omega_2, a \in \mathbb{R}$ with $a \neq \frac{1}{2}$ and $\omega_2 \neq 2$. For $k = 0, 1, 2, ...$ until the iteration sequence $\{(x^{(k)^T}, y^{(k)^T})^T\}$ is convergent, compute*

$$y^{(k+1)} = y^{(k)} + \frac{1}{(1-a\omega_2)[1-(1-a)\omega_2]} Q^{-1} \left\{ B^T[(\tau_2 - \tau_1\omega_2)x^{(k)} + \tau_1\omega_2 A^{-1}(b_1 - By^{(k)})] - \tau_2 b_2 \right\}$$
$$x^{(k+1)} = (1 - \tau_1)x^{(k)} + A^{-1} \left\{ B\left[(\omega_1 - \tau_1)y^{(k)} - \omega_1 y^{(k+1)}\right] + \tau_1 b_1 \right\},$$

where Q is an approximation of the Schur complement matrix $B^T A^{-1} B$.

Note that in the above algorithm we first compute $y^{(k+1)}$ and then $x^{(k+1)}$, whereas in the GMESOR method we had the reverse computations. If $\tau = \tau_1 = \tau_2$ and $\omega = \omega_1 = \omega_2$ we have the GPSD method.
If $\omega_2 = 0$ then the algorithmic form of the GMPSD method simplifies to

$$y^{(k+1)} = y^{(k)} + \tau_2 Q^{-1}(B^T x^{(k)} - b_2)$$
$$x^{(k+1)} = (1 - \tau_1)x^{(k)} + \tau_1 A^{-1}(b_1 - B y^{(k+1)}) \qquad (122)$$

The above form is the same as that of the GSOR method if we use $\mathcal{D} - \Omega\mathcal{U}$ instead of $\mathcal{D} - \Omega\mathcal{L}$ as the preconditioned matrix in the GSOR method and will

be referred as the simplified GMPSD method. In the following theorem we find the functional relationship for the GMPSD method between the eigenvalues λ of the iteration matrix $\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2, a)$ and the eigenvalues μ of the matrix J.

Theorem 3.1 *Let $A \in \mathbb{R}^{m \times m}$ be symmetric positive definite, $B \in \mathbb{R}^{m \times n}$ be of full column rank and $Q \in \mathbb{R}^{n \times n}$ be nonsingular and symmetric. If $\lambda \neq 1 - \tau_1$ is an eigenvalue of the matrix $\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2, a)$ and if μ satisfies*

$$\lambda^2 + \lambda \left(\tau_1 - 2 + \frac{\tau_1 \omega_2 + \tau_2 \omega_1 - \tau_1 \omega_1 \omega_2}{(1 - a\omega_2)[1 - (1 - a)\omega_2]} \mu \right) + 1 - \tau_1 + \frac{\tau_1 \tau_2 - \tau_1 \omega_2 - \tau_2 \omega_1 + \tau_1 \omega_1 \omega_2}{(1 - a\omega_2)[1 - (1 - a)\omega_2]} \mu = 0,$$
(123)

where $a \neq \frac{1}{2}$ and $\omega_2 \neq 2$, then μ is an eigenvalue of the matrix $J = Q^{-1} B^T A^{-1} B$. Conversely, if μ is an eigenvalue of J and if $\lambda \neq 1 - \tau_1$ satisfies (123), then λ is an eigenvalue of $\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2, a)$. In addition, $\lambda = 1 - \tau_1$ is an eigenvalue of $\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2, a)$ (if $m > n$) with the corresponding eigenvector $(x^T, 0)^T$, where $x \in \mathcal{N}(B^T)$.

Proof Clearly, the eigenvalues μ of the matrix $J = Q^{-1} B^T A^{-1} B$ are real and non-zero. Let λ be a nonzero eigenvalue of the iteration matrix $\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2, a)$ defined in (117), and $(x, y)^T \in \mathbb{R}^{m+n}$ be the corresponding eigenvector. Then, we have that

$$\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2, a) \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}$$

or because of (117)

$$[(\mathcal{D} - \Omega\mathcal{L})\mathcal{D}^{-1}(\mathcal{D} - \Omega\mathcal{U}) - T\mathcal{A}] \begin{pmatrix} x \\ y \end{pmatrix} = \lambda(\mathcal{D} - \Omega\mathcal{L})\mathcal{D}^{-1}(\mathcal{D} - \Omega\mathcal{U}) \begin{pmatrix} x \\ y \end{pmatrix}. \quad (124)$$

From (124), because of (4), we have that

$$\begin{pmatrix} (1 - \tau_1)A & (\omega_1 - \tau_1)B \\ (\tau_2 - \omega_2)B^T & (1 - a\omega_2)[1 - (1 - a)\omega_2]Q - \omega_1\omega_2 B^T A^{-1} B \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$
$$= \lambda \begin{pmatrix} A & \omega_1 B \\ -\omega_2 B^T & (1 - a\omega_2)[1 - (1 - a)\omega_2]Q - \omega_1\omega_2 B^T A^{-1} B \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Decoupling we have that

$$\begin{cases} (1 - \tau_1)Ax + (\omega_1 - \tau_1)By = \lambda Ax + \lambda\omega_1 By \\ (\tau_2 - \omega_2)B^T x + \{(1 - a\omega_2)[1 - (1 - a)\omega_2]Q - \omega_1\omega_2 B^T A^{-1} B\}y \\ \qquad = -\lambda\omega_2 B^T x + \lambda\{(1 - a\omega_2)[1 - (1 - a)\omega_2]Q - \omega_1\omega_2 B^T A^{-1} B\}y \end{cases}$$

or equivalently

$$\begin{cases} (1 - \tau_1 - \lambda)x = [(\lambda - 1)\omega_1 + \tau_1]A^{-1}By \\ (\tau_2 - \omega_2 + \lambda\omega_2)Q^{-1}B^T x = (\lambda - 1)\{(1 - a\omega_2)[1 - (1 - a)\omega_2]I - \omega_1\omega_2 J\}y. \end{cases}$$
(125)

From the first equality in (125) we get

$$(1 - \tau_1 - \lambda)Q^{-1}B^T x = [(\lambda - 1)\omega_1 + \tau_1]Jy,$$

and hence, when $\lambda \neq 1 - \tau_1$,

$$Q^{-1}B^T x = \frac{(\lambda - 1)\omega_1 + \tau_1}{1 - \tau_1 - \lambda}Jy. \tag{126}$$

It then follows from (126) and the second equality in (125) that

$$(\lambda - 1)(1 - a\omega_2)[1 - (1 - a)\omega_2](1 - \tau_1 - \lambda)y$$
$$= \{[(\lambda - 1)\omega_2 + \tau_2][(\lambda - 1)\omega_1 + \tau_1] + (\lambda - 1)(1 - \tau_1 - \lambda)\omega_1\omega_2\}Jy.$$

If $\lambda = 1 - \tau_1 \neq 0$, then from the first and the second equality of (125) we have, respectively, $By = 0$ and $\tau_1\{(1 - a\omega_2)[1 - (1 - a)\omega_2]Q - \omega_1\omega_2 B^T A^{-1}B\}y = (\tau_1\omega_2 - \tau_2)B^T x$. It then follows that $y = 0$ and $x \in \mathcal{N}(B^T)$, where $\mathcal{N}(B^T)$ is the null space of the matrix B^T. Hence, $\lambda = 1 - \tau_1$ is an eigenvalue of $\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2, a)$ with the corresponding eigenvector $(x^T, 0)^T$, where $x \in \mathcal{N}(B^T)$. Therefore, the eigenvalues λ (except for $\lambda = 1 - \tau_1$) of the matrix $\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2, a)$ and the eigenvalues μ of the matrix J satisfy the functional relationship

$$(\lambda - 1)(1 - a\omega_2)[1 - (1 - a)\omega_2](1 - \tau_1 - \lambda)$$
$$= \{[(\lambda - 1)\omega_2 + \tau_2][(\lambda - 1)\omega_1 + \tau_1] + (\lambda - 1)(1 - \tau_1 - \lambda)\omega_1\omega_2\}\mu.$$

This means that λ satisfies the quadratic equation (123). $\qquad\square$

Corollary 3.1 *Let $A \in \mathbb{R}^{m \times m}$ be symmetric positive definite, $B \in \mathbb{R}^{m \times n}$ be of full column rank and $Q \in \mathbb{R}^{n \times n}$ be nonsingular and symmetric.*
1. The nonzero eigenvalues of the iteration matrix $\mathcal{G}(\tau, \omega_1, \omega_2, a)$ of the GMPSD(3) method are given by $\lambda = 1 - \tau$ or if $a \neq \frac{1}{2}$ and $\omega_2 \neq 2$ by

$$\lambda^2 + \lambda\left(\tau - 2 + \frac{\tau\hat{\omega}}{(1 - a\omega_2)[1 - (1 - a)\omega_2]}\mu\right) + 1 - \tau + \frac{\tau(\tau - \hat{\omega})}{(1 - a\omega_2)[1 - (1 - a)\omega_2]}\mu = 0 \tag{127}$$

where

$$\hat{\omega} = \omega_1 + \omega_2 - \omega_1\omega_2. \tag{128}$$

2. The nonzero eigenvalues of the iteration matrix $\mathcal{S}(\omega_1, \omega_2, a)$ of the GMSSOR method are given by $\lambda = 1 - \hat{\omega}$ or if $a \neq \frac{1}{2}$ and $\omega_2 \neq 2$ by

$$\lambda^2 + \lambda\left(\hat{\omega} - 2 + \frac{\hat{\omega}^2}{(1 - a\omega_2)[1 - (1 - a)\omega_2]}\mu\right) + 1 - \hat{\omega} = 0 \tag{129}$$

where $\hat{\omega}$ is given by (128).
The nonzero eigenvalues of the iteration matrix $\mathcal{G}(\tau, \omega, a)$ of the GPSD method are given by $\lambda = 1 - \tau$ or if $a \neq \frac{1}{2}$ and $\omega \neq 2$ by

$$\lambda^2 + \lambda\left(\tau - 2 + \frac{\tau\hat{\omega}}{(1 - a\omega)[1 - (1 - a)\omega]}\mu\right) + 1 - \tau + \frac{\tau(\tau - \hat{\omega})}{(1 - a\omega)[1 - (1 - a)\omega]}\mu = 0 \tag{130}$$

where now

$$\hat{\omega} = \omega(2 - \omega). \qquad (131)$$

and

$$a\omega \neq 1 \ \ and \ \ (1 - a)\omega \neq 1 \qquad (132)$$

4. The nonzero eigenvalues of the iteration matrix $\mathcal{S}(\omega, a)$ of the GSSOR method are given by $\lambda = 1 - \hat{\omega}$ or if $a \neq \frac{1}{2}$ and $\omega \neq 2$ by

$$\lambda^2 + \lambda \left(\hat{\omega} - 2 + \frac{\hat{\omega}^2}{(1 - a\omega)[1 - (1 - a)\omega]} \mu \right) + 1 - \hat{\omega} = 0 \qquad (133)$$

where $\hat{\omega}$ is given by (131).

Proof The iteration matrix $\mathcal{G}(\tau, \omega_1, \omega_2, a)$ of the GMPSD(3) is obtained by letting $\tau = \tau_1 = \tau_2$ in $\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2, a)$ given by (117). Using the matrix $\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2, a)$ and following a similar approach as in the proof of Theorem 3.1 we find the functional relationship (127). Similarly, we find (129), (130) and (133). $\qquad \square$

3.2 Convergence

If the matrix Q is positive definite and $a = 0$ sufficient conditions for the GMPSD method to converge are given by the following theorem.

Theorem 3.2 *Consider the GMPSD method. Let $A \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m \times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1}B$ by μ_{min} and μ_{max}, respectively. Then, $\rho(\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2)) < 1$ if the parameters τ_1, τ_2, ω_1 and ω_2 lie in the region defined in the cases of Table 5 with $0 < \tau_1 < 2$ and*

$$\begin{aligned}
\omega_{11}^*(\mu) &= \frac{\tau_1(2\omega_2 - \tau_2)}{2(\tau_1\omega_2 - \tau_2)} + \frac{(\tau_1 - 2)(1 - \omega_2)}{\tau_1\omega_2 - \tau_2}\frac{1}{\mu}, \qquad \omega_{21}^* = \frac{\tau_2}{\tau_1}, \\
\omega_{12}^*(\mu) &= \frac{\tau_1(\omega_2 - \tau_2)}{\tau_1\omega_2 - \tau_2} + \frac{\tau_1(1 - \omega_2)}{\tau_1\omega_2 - \tau_2}\frac{1}{\mu}, \qquad \omega_{22}^*(\mu) = 1 - \frac{\tau_1\tau_2\mu}{4}.
\end{aligned} \qquad (134)$$

Table 5 Sufficient conditions for the GMPSD method to converge.

Cases	ω_2 − Domain	ω_1 − Domain	τ_2 − Domain
1	$\omega_{21}^* < \omega_2 < \omega_{22}^*(\mu_{max})$	$\omega_{11}^*(\mu_{max}) < \omega_1 < \omega_{12}^*(\mu_{min})$	
2	$\omega_2 < \omega_{21}^*$		$0 < \tau_2 < \dfrac{4\tau_1}{4 + \tau_1^2\mu_{max}}$
3	$\omega_2 < \omega_{22}^*(\mu_{max})$	$\omega_{12}^*(\mu_{min}) < \omega_1 < \omega_{11}^*(\mu_{max})$	$\dfrac{4\tau_1}{4 + \tau_1^2\mu_{min}} < \tau_2$
4	$1 < \omega_2 < \omega_{22}^*(\mu_{min})$		$\tau_2 < 0$

Proof Recall that $\lambda = 1 - \tau_1 \neq 0$ is an eigenvalue of $\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2)$ and if $\lambda \neq 1 - \tau_1$ then the eigenvalues of $\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2)$ are given by (123). If $\lambda = 1 - \tau_1 \neq 0$, then the GMPSD method is convergent if and only if $|\lambda| < 1$, that is $|1 - \tau_1| < 1$, or

$$0 < \tau_1 < 2. \tag{135}$$

If $\lambda \neq 1 - \tau_1$, then (123) holds and by Lemma 2.1 page 171 of [55], it follows that the GMPSD method is convergent if and only if

$$|c| < 1 \quad \text{and} \quad |b| < 1 + c \tag{136}$$

where

$$c = 1 - \tau_1 + \frac{\tau_1 \omega_1 \omega_2 - \tau_1 \omega_2 - \tau_2 \omega_1 + \tau_1 \tau_2}{1 - \omega_2} \mu \tag{137}$$

and

$$b = 1 + c - \frac{\tau_1 \tau_2}{1 - \omega_2} \mu. \tag{138}$$

From the first inequality of (136) it follows that

$$0 < 1 + c < 2. \tag{139}$$

From the second inequality of (136), because of (138), we have

$$0 < \frac{\tau_1 \tau_2 \mu}{2(1 - \omega_2)} < 1 + c. \tag{140}$$

Combining (139) and (140) it follows that

$$0 < \frac{\tau_1 \tau_2 \mu}{2(1 - \omega_2)} < 1 + c < 2. \tag{141}$$

In order for (141) to hold we must have that

$$0 < \frac{\tau_1 \tau_2 \mu}{2(1 - \omega_2)} < 2,$$

or because of (135)

$$0 < \frac{\tau_2}{1 - \omega_2} < \frac{4}{\tau_1 \mu}. \tag{142}$$

Inequalities (141), because of (137), become

$$\frac{\tau_1 (2\omega_2 - \tau_2)\mu}{2(1 - \omega_2)} + \tau_1 - 2 < \omega_1 \frac{\tau_1 \omega_2 - \tau_2}{1 - \omega_2} \mu < \tau_1 + \frac{\tau_1 (\omega_2 - \tau_2)\mu}{1 - \omega_2}. \tag{143}$$

In the sequel we distinguish the following two cases to study (143). Case I: $\tau_2 > 0$ and $1 - \omega_2 > 0$ and Case II: $\tau_2 < 0$ and $1 - \omega_2 < 0$. In addition, we distinguish the following two subcases for each of the above cases. (i): $\tau_1 \omega_2 - \tau_2 > 0$ and (ii): $\tau_1 \omega_2 - \tau_2 < 0$. Next, we will study only the subcase (i) of Case I, since the other cases can be treated similarly. For this case, we have that

$$\frac{\tau_2}{\tau_1} < \omega_2 < 1, \quad \text{if} \quad 0 < \tau_2 < \tau_1 \tag{144}$$

and from the second part of (142)

$$\omega_2 < 1 - \frac{\tau_1 \tau_2 \mu}{4}.$$ (145)

From (144) and (145) it follows that

$$\frac{\tau_2}{\tau_1} < \omega_2 < \min\left\{1, 1 - \frac{\tau_1 \tau_2 \mu}{4}\right\}, \quad 0 < \tau_2 < \tau_1$$

or

$$\frac{\tau_2}{\tau_1} < \omega_2 < 1 - \frac{\tau_1 \tau_2 \mu}{4}, \quad 0 < \tau_2 < \tau_1$$ (146)

which holds if $\frac{\tau_2}{\tau_1} < 1 - \frac{\tau_1 \tau_2 \mu}{4}$. Therefore, we have that (146) holds if

$$\omega_{21}^* < \omega_2 < \omega_{22}^*(\mu), \quad 0 < \tau_2 < \frac{4\tau_1}{4 + \tau_1^2 \mu}$$ (147)

where $\omega_{21}^*, \omega_{22}^*(\mu)$ are given by (134). Furthermore, from (143), we have that

$$\omega_{11}^*(\mu) < \omega_1 < \omega_{12}^*(\mu)$$ (148)

where $\omega_{11}^*(\mu), \omega_{12}^*(\mu)$ are given by (134). Studying the monotonicity of $\omega_{22}^*(\mu), \omega_{11}^*(\mu)$ and $\omega_{12}^*(\mu)$ with respect to μ we have that $sign\frac{\partial \omega_{22}^*(\mu)}{\partial \mu} = -1, sign\frac{\partial \omega_{11}^*(\mu)}{\partial \mu} = +1$ and $sign\frac{\partial \omega_{12}^*(\mu)}{\partial \mu} = +1$. Hence, case 1 of Table 5 is proved. Treating similarly subcase (ii) of Case I and subcases (i) and (ii) of Case II, we can prove the rest of the cases in Table 5. □

The convergence conditions for the GMPSD(3) are given by the following.

Corollary 3.2 *Consider the GMPSD(3) method. Let $A \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m \times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1} B^T A^{-1} B$ by μ_{min} and μ_{max}, respectively. Then, $\rho(\mathcal{G}(\tau, \omega_1, \omega_2)) < 1$ if*

$$0 < \tau < 2, \quad \omega_2 < \omega_2^*(\mu_{max}) \quad and \quad \omega_{13}^*(\mu_{max}) < \omega_1 < \omega_{14}^*(\mu_{max})$$ (149)

where

$$\omega_2^*(\mu) = 1 - \frac{\tau^2 \mu}{4}, \quad \omega_{13}^*(\mu) = \frac{\tau - \omega_2}{1 - \omega_2} - \frac{1}{\mu}, \quad \omega_{14}^*(\mu) = \frac{2 - \tau}{\tau \mu} + \frac{\tau - 2\omega_2}{2(1 - \omega_2)}.$$ (150)

Proof Letting $a = 0$ in the functional relationship (127) and following a similar approach as in the proof of Theorem 3.2, we can prove (149). □

Note that analogous results hold when $Q \in \mathbb{R}^{n \times n}$ is symmetric negative definite.

3.3 Optimum parameters

In the following theorem the optimum parameters of the GMPSD method are determined assuming that the matrix Q is symmetric positive definite and $a \neq 0$.

Theorem 3.3 *Consider the GMPSD method. Let $A \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m \times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1}B$ by μ_{min} and μ_{max}, respectively. Then the spectral radius of the GMPSD method, $\rho(\mathcal{G}(\tau_1, \tau_2, \omega_1, \omega_2, a))$, is minimized for any $\omega_2 \neq \frac{\tau_{2_{opt}}}{\tau_{1_{opt}}}$ at*

$$\omega_{1_{opt}} = \frac{\tau_{1_{opt}}(\tau_{2_{opt}} - \omega_2)}{\tau_{2_{opt}} - \tau_{1_{opt}}\omega_2}, \tag{151}$$

$$\tau_{1_{opt}} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2} \quad and \quad \tau_{2_{opt}} = \frac{(1 - a\omega_2)[1 - (1-a)\omega_2]}{\sqrt{\mu_{min}\mu_{max}}} \tag{152}$$

and its corresponding value is

$$\rho(\mathcal{G}(\tau_{1_{opt}}, \tau_{2_{opt}}, \omega_{1_{opt}}, \omega_2, a)) = \frac{\sqrt{\mu_{max}} - \sqrt{\mu_{min}}}{\sqrt{\mu_{max}} + \sqrt{\mu_{min}}}. \tag{153}$$

Proof Following a similar approach as in Theorem 2.10, using the functional relationship (123) and requiring $|\tilde{\lambda}_1| = |\tilde{\lambda}_N|$ we find

$$\tau_1\omega_1\omega_2 - \tau_2\omega_1 - \tau_1\omega_2 + \tau_1\tau_2 = 0. \tag{154}$$

Therefore, (123) because of (154), becomes

$$\lambda^2 + \lambda(\tau_1 - 2 + \tau_1\hat{\tau}_2\mu) + 1 - \tau_1 = 0 \tag{155}$$

with

$$\hat{\tau}_2 = \frac{\tau_2}{(1 - a\omega_2)[1 - (1-a)\omega_2]} \tag{156}$$

which is the functional relationship of the GSOR method [8] with the only difference that now we have $(1 - a\omega_2)[1 - (1-a)\omega_2]$ instead of $1 - a\omega_2$ in the denominator of $\hat{\tau}_2$ (see (103)), hence (151), follows from (154) whereas (152) and (153) hold because of (155), (156) and Theorem 4.1 in [8].   □

Corollary 3.3 *Consider the simplified GMPSD method. Let $A \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m \times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1}B$ by μ_{min} and μ_{max}, respectively. Then the spectral radius of the simplified GMPSD method, $\rho(\mathcal{G}(\tau_1, \tau_2, \omega_1, 0, 0))$, is minimized at*

$$\omega_{1_{opt}} = \tau_{1_{opt}}, \tag{157}$$

$$\tau_{1_{opt}} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2} \quad and \quad \tau_{2_{opt}} = \frac{1}{\sqrt{\mu_{min}\mu_{max}}} \tag{158}$$

and its corresponding value is

$$\rho(\mathcal{G}(\tau_{1_{opt}}, \tau_{2_{opt}}, \omega_{1_{opt}}, 0, 0)) = \frac{\sqrt{\mu_{max}} - \sqrt{\mu_{min}}}{\sqrt{\mu_{max}} + \sqrt{\mu_{min}}}. \tag{159}$$

Proof Letting $\omega_2 = 0$, (151), (152) and (153) yield (157), (158) and (159), respectively. □

It is worth noting here that the optimum values of $\tau_{1_{opt}}$ and $\tau_{2_{opt}}$ of the simplified GMPSD method are identical to the optimum values of $\omega_{1_{opt}}$ and $\omega_{2_{opt}}$ of the GSOR method, respectively.

Corollary 3.4 *Consider the GMPSD(3) method. Let $A \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m \times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1}B$ by μ_{min} and μ_{max}, respectively. If $\mu_{max} < \frac{1}{4}$ or if $\mu_{max} > \frac{1}{4}$ and either (i) $\mu_{min} < \mu^*$ or (ii) $\mu_{min} \geq \mu^*$ and $a_1 \leq a \leq a_2$, then the spectral radius of the GMPSD(3) method, $\rho(\mathcal{G}(\tau, \omega_1, \omega_2, a))$, is minimized at*

$$\omega_{1_{opt}} = \frac{\tau_{opt} - \omega_{2_{opt}}}{1 - \omega_{2_{opt}}}, \tag{160}$$

$$\tau_{opt} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2} \quad and \quad \omega_{2_{opt}} = \frac{\sigma}{2\left[1 \pm \sqrt{1 - a(1-a)\sigma}\right]} \tag{161}$$

and its corresponding value is

$$\rho(\mathcal{G}(\tau_{opt}, \omega_{1_{opt}}, \omega_{2_{opt}}, a)) = \frac{\sqrt{\mu_{max}} - \sqrt{\mu_{min}}}{\sqrt{\mu_{max}} + \sqrt{\mu_{min}}}. \tag{162}$$

where

$$\mu^* = \frac{\mu_{max}}{(1 - 2\sqrt{\mu_{max}})^2}, \quad a_1 = \frac{2}{\sigma - \sqrt{\sigma(\sigma - 4)}}, \quad a_2 = \frac{2}{\sigma + \sqrt{\sigma(\sigma - 4)}} \tag{163}$$

with

$$\sigma = 4(1 - M) \quad and \quad M = \frac{4\mu_{min}\mu_{max}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2}. \tag{164}$$

Proof Recall that GMPSD(3) is obtained by setting $\tau_1 = \tau_2$ in GMPSD. Therefore, (160), (161) and (162) are obtained by (151), (152) and (153) respectively. In particular, by letting $\tau_{1_{opt}} = \tau_{2_{opt}}$ it follows from (152) that

$$a(1-a)\omega_2^2 - \omega_2 + 1 - M = 0, \tag{165}$$

where M is given by (176). This quadratic has real roots when

$$a^2\sigma - a\sigma + 1 \geq 0, \tag{166}$$

where σ is given by (176). Considering (178) as a quadratic we distinguish two cases. Case 1: $\Delta_a < 0$, Case 2: $\Delta_a \geq 0$ where $\Delta_a = \sigma(\sigma - 4)$.

Case 1: $\Delta_a < 0$. In this case we require $\sigma > 0$ since $\sigma - 4 < 0$ or in view of (176)

$$\sqrt{\mu_{min}}(1 - 2\sqrt{\mu_{max}}) > -\sqrt{\mu_{max}}. \tag{167}$$

But, (179) holds if either $\mu_{max} < \frac{1}{4}$ or if $\mu_{max} > \frac{1}{4}$ and $\mu_{min} < \mu^*$ and (i) is proved.

Case 2: $\Delta_a \geq 0$. In this case we require $\sigma \leq 0$ since $\sigma - 4 < 0$ or, because of (176),

$$\sqrt{\mu_{min}}(1 - 2\sqrt{\mu_{max}}) \leq -\sqrt{\mu_{max}} \tag{168}$$

which holds if $\mu_{max} > \frac{1}{4}$ and

$$\mu_{min} \geq \mu^*. \tag{169}$$

In this case, for (178) to hold, a must lie in the range given by (ii). Hence, the proof of the theorem is complete.   □

Corollary 3.5 *Consider the GMSSOR method. Let $A \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m \times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1}B$ by μ_{min} and μ_{max}, respectively. If $\mu_{max} < \frac{1}{4}$ or if $\mu_{max} > \frac{1}{4}$ and either (i) $\mu_{min} < \mu^*$ or (ii) $\mu_{min} \geq \mu^*$ and $a_1 \leq a \leq a_2$, then the spectral radius of the GMSSOR method, $\rho(\mathcal{G}(\omega_1, \omega_2, a))$, is minimized at*

$$\omega_{1_{opt}} = \frac{\hat{\omega}_{opt} - \omega_{2_{opt}}}{1 - \omega_{2_{opt}}} \tag{170}$$

where

$$\hat{\omega}_{opt} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2} \quad and \quad \omega_{2_{opt}} = \frac{\sigma}{2\left[1 \pm \sqrt{1 - a(1-a)\sigma}\right]} \tag{171}$$

and its corresponding value is

$$\rho(\mathcal{G}(\omega_{1_{opt}}, \omega_{2_{opt}}, a)) = \frac{\sqrt{\mu_{max}} - \sqrt{\mu_{min}}}{\sqrt{\mu_{max}} + \sqrt{\mu_{min}}}. \tag{172}$$

where μ^*, a_1, a_2, σ are given by (175), (176).

Proof Recall that GMSSOR is obtained by setting $\tau_1 = \tau_2 = \hat{\omega}$ in GMPSD. Therefore, (170), (171) and (172) are obtained by (151), (152) and (153), respectively. Indeed, as in GMPSD(3), since $\tau_{1_{opt}} = \tau_{2_{opt}}$ it follows that (177) holds also and by the analysis of the proof of Corollary 3.4, we have that (170), (171) and (172) hold under the same conditions as in Corollary 3.4.   □

Corollary 3.6 *Consider the GPSD method. Let $A \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m \times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^T A^{-1}B$ by μ_{min} and μ_{max}, respectively. If $\mu_{max} < \frac{1}{4}$ or if $\mu_{max} > \frac{1}{4}$ and either (i)*

$\mu_{min} < \mu^*$ or (ii) $\mu_{min} \geq \mu^*$ and $a_1 \leq a \leq a_2$, then the spectral radius of the GPSD method, $\rho(\mathcal{G}(\tau, \omega, a))$, is minimized at

$$\tau_{opt} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2} \quad and \quad \omega_{opt} = \frac{\sigma}{2\left[1 \pm \sqrt{1 - a(1-a)\sigma}\right]} \tag{173}$$

and its corresponding value is

$$\rho(\mathcal{G}(\tau_{opt}, \omega_{opt}, a)) = \frac{\sqrt{\mu_{max}} - \sqrt{\mu_{min}}}{\sqrt{\mu_{max}} + \sqrt{\mu_{min}}} \tag{174}$$

where

$$\mu^* = \frac{\mu_{max}}{(1 - 2\sqrt{\mu_{max}})^2}, \quad a_1 = \frac{2}{\sigma - \sqrt{\sigma(\sigma - 4)}}, \quad a_2 = \frac{2}{\sigma + \sqrt{\sigma(\sigma - 4)}} \tag{175}$$

with

$$\sigma = 4(1 - M) \quad and \quad M = \frac{4\mu_{min}\mu_{max}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2}. \tag{176}$$

Proof GPSD follows from GMPSD by letting $\tau = \tau_1 = \tau_2$ and $\omega = \omega_1 = \omega_2$ or $\tau_{opt} = \tau_{1_{opt}} = \tau_{2_{opt}}$ and $\omega_{opt} = \omega_{1_{opt}} = \omega_{2_{opt}}$. By equating the expressions of $\tau_{1_{opt}}$ and $\tau_{2_{opt}}$ given by (152) we obtain

$$a(1-a)\omega_2^2 - \omega_2 + 1 - M = 0, \tag{177}$$

where M is given by (176). This quadratic has real roots when

$$a^2\sigma - a\sigma + 1 \geq 0, \tag{178}$$

where σ is given by (176). We distinguish two cases. Case 1: $\Delta_a < 0$, Case 2: $\Delta_a \geq 0$ where $\Delta_a = \sigma(\sigma - 4)$.

Case 1: $\Delta_a < 0$. In this case we require $\sigma > 0$ since $\sigma - 4 < 0$ or in view of (176)

$$\sqrt{\mu_{min}}(1 - 2\sqrt{\mu_{max}}) > -\sqrt{\mu_{max}}. \tag{179}$$

But, (179) holds if either $\mu_{max} < \frac{1}{4}$ or if $\mu_{max} > \frac{1}{4}$ and $\mu_{min} < \mu^*$ hence (i) is proved.

Case 2: $\Delta_a \geq 0$. In this case we require $\sigma \leq 0$ since $\sigma - 4 < 0$ or, because of (176),

$$\sqrt{\mu_{min}}(1 - 2\sqrt{\mu_{max}}) \leq -\sqrt{\mu_{max}} \tag{180}$$

which holds if $\mu_{max} > \frac{1}{4}$ and

$$\mu_{min} \geq \mu^*. \tag{181}$$

In this case, for (178) to hold, a must lie in the range given by (ii). Therefore, it follows that ω_{opt} is given by (173). $\qquad\square$

Analogous results hold in case where the matrix Q is symmetric negative definite.

4 Numerical results

In this section we study the numerical solution of the following linear Stokes equation

$$\begin{cases} -\mu \Delta \mathbf{u} + \nabla w = \tilde{f}, & \Omega \\ \nabla \cdot \mathbf{u} = \tilde{g}, & \Omega \\ \mathbf{u} = 0, & \partial\Omega \\ \int_{\Omega} w(x)dx = 0, \end{cases} \tag{182}$$

where $\Omega = (0,1) \times (0,1) \subset \mathbb{R}^2$, $\partial\Omega$ is the boundary of Ω, Δ is the componentwise Laplace operator, $\mathbf{u}$ is a vector-valued function representing the velocity and w is a scalar function representing the pressure. Furthermore, we assume that the functions $\tilde{f}, \tilde{g}$ are constant. By discretizing (182) with the upwind scheme, we obtain the system of linear equations (1), in which [7]

$$A = \begin{pmatrix} I \otimes T + T \otimes I & 0 \\ 0 & I \otimes T + T \otimes I \end{pmatrix} \in \mathbb{R}^{2p^2 \times 2p^2},$$

$$B = \begin{pmatrix} I \otimes F \\ F \otimes I \end{pmatrix} \in \mathbb{R}^{2p^2 \times p^2}$$

with

$$T = \frac{\mu}{h^2} \cdot tridiag(-1, 2, -1) \in \mathbb{R}^{p \times p}, \quad F = \frac{1}{h} \cdot tridiag(-1, 1, 0) \in \mathbb{R}^{p \times p},$$

$h = \frac{1}{p+1}$ being the discretization mesh size and $\otimes$ the Kronecker product symbol. For this example, we let $\mu = 1$, $m = 2p^2$ and $n = p^2$. Hence, the total number of variables is $m + n = 3p^2$.

We choose the matrix Q to be an approximation to $B^T A^{-1} B$. The reason being that if $Q \simeq B^T A^{-1} B$ then $J = Q^{-1} B^T A^{-1} B \simeq I$. In this case the ratio of the maximum to the minimum eigenvalue of the matrix J becomes minimum and its value is approximately 1. As a consequence, the spectral radius of the iteration matrix of the GMESOR and GMPSD methods attains its minimum value. We choose Q, according to the following two cases:

1. $Q = B^T \hat{A}^{-1} B$, $\hat{A} = \text{tridiag}(A)$
2. $Q = B^T \hat{A}^{-1} B$, $\hat{A} = \text{diag}(A)$,

where $\hat{A}$ is the tridiagonal or the diagonal part of A. The choice of the matrix $\hat{A}$ instead of A is due to the difficulty in computing the inverse matrix of A. In this example the eigenvalues of Q are real and positive. In actual computations, we choose the right-hand-side vector $(b^T, q^T)^T \in \mathbb{R}^{m+n}$ such that the exact solution of the augmented linear system (1) is $((x^*)^T, (y^*)^T)^T = (1, 1, ..., 1)^T \in \mathbb{R}^{m+n}$, and perform all runs in MATLAB (version R_2012b) with a machine precision 10^{-16}. The machine used was an Intel i5 personal computer with 6G memory. In our computations, all runs are started from the

initial vector $\left((x^{(0)})^T, y^{(0)T}\right)^T = 0$, and terminated if the current iterations satisfy

$$RES = \frac{\sqrt{||b - Ax^{(k)} - By^{(k)}||_2^2 + ||q - B^T x^{(k)}||_2^2}}{\sqrt{||b - Ax^{(0)} - By^{(0)}||_2^2 + ||q - B^T x^{(0)}||_2^2}} \le 10^{-9},$$

where RES is the norm of absolute residual vectors, or if the numbers of the prescribed iterations $k_{max} = 1200$ are exceeded. We also use the same example to compare our methods with the PHSS [7] and Krylov subspace methods [41], [42], [49].

In Table 6 we computed the optimal parameters $\tau_{1_{opt}}, \tau_{2_{opt}}$ and $\omega_{2_{opt}}$ and the optimal spectral radius ρ_{opt} of the GMESOR method, for various problem sizes (m,n) using (91), (92) and (93). Furthermore, we computed the optimum parameters $\tau_{2_{opt}}(exp)$, $\omega_{2_{opt}}(exp)$ and the spectral radii $\rho(\tau_{2_{opt}}(exp))$ and $\rho(\omega_{2_{opt}}(exp))$, experimentally by trial and error. The parameter τ_1 was kept fixed and was given its optimum value. Our results show that $\rho_{opt} \simeq \rho(\tau_{2_{opt}}(exp)) \simeq \rho(\omega_{2_{opt}}(exp))$ and $\omega_{2_{opt}} = \tau_{2_{opt}} \simeq \tau_{2_{opt}}(exp) \simeq \omega_{2_{opt}}(exp)$ thus verifying Theorem 2.10. The numerical results in Table 7 verify that the parameter a may be chosen arbitrary, while the minimum value of $\rho(\mathcal{H}(\tau_1, \tau_2, \omega_2, a))$ remains approximately the same. $\rho(\mathcal{H}(\tau_{1_{opt}}, \tau_{2_{opt}}, \omega_{2_{opt}}, a))$ was computed using Matlab. The slightly different values are due to rounding errors. Finally, in Table 8 we list numerical results with respect to the number of total iteration steps (denoted by "ITER"), the elapsed CPU time in seconds (denoted by "CPU") and RES for the GSOR, GMESOR and Simplified GMPSD iterative methods. We remark that our numerical results verify the validity of theorem 2.10 and corollary 3.3, since GSOR, GMESOR and Simplified GMPSD methods require the same number of iterations for convergence. Indeed, this was expected since all these methods have the same spectral radius for the optimum values of their parameters. Note that all the aforementioned methods require approximately the same computing time. Furthermore, for comparison purposes we also considered the PHSS(a^*), GMRES, GMRES($\#$), PGMRES and PGMRES($\#$) methods. The integer $\#$ in GMRES($\#$) and PGMRES($\#$) methods denotes the number of restarting steps, while the integer a^* denotes the theoretical optimal parameter of the PHSS method. We also list numerical results with respect to the number of total iteration steps and the elapsed CPU time in seconds for these methods. The preconditioned matrix Q in PHSS(a^*) is given by the aforementioned cases 1 and 2. The preconditioner, say K, for the PGMRES and PGMRES($\#$) methods is given by [17], [18], [39], [54]

$$K = \begin{bmatrix} \hat{A} & 0 \\ 0 & I \end{bmatrix}.$$

We remark that the GSOR, GMESOR and Simplified GMPSD methods always outperform the other testing methods, except of the PHSS(a^*) method, considerably with respect to iteration steps as p increases. However, the overall computing time of the GSOR, GMESOR and Simplified GMPSD methods is

much smaller than that of all the other testing methods. With * we denote that the method converges but after too many hours. With regard to the matrix Q, Case 1 is the best choice for all methods tested as it requires the least iteration steps and CPU times.

Table 6 Experimental results showing that $\omega_{2_{opt}} = \tau_{2_{opt}}$ in GMESOR.

m		128	512	1152
n		64	256	576
m+n		192	768	1728
Case 1	$\tau_{1_{opt}}$	0.663309	0.442911	0.330674
	$\tau_{2_{opt}}$	0.499375	0.285422	0.198468
	$\omega_{2_{opt}}$	0.499375	0.285422	0.198468
	ρ_{opt}	0.580251	0.746384	0.811229
	$\tau_{2_{opt}}(exp)$	0.5	0.286	0.199
	$\rho(\tau_{2_{opt}}(exp))$	0.582936	0.750508	0.823517
	$\omega_{2_{opt}}(exp)$	0.499	0.285	0.198
	$\rho(\omega_{2_{opt}}(exp))$	0.581866	0.749401	0.822877
Case 2	$\tau_{1_{opt}}$	0.757767	0.631420	0.558518
	$\tau_{2_{opt}}$	1.950825	2.529944	2.974309
	$\omega_{2_{opt}}$	1.950825	2.529944	2.974309
	ρ_{opt}	0.492171	0.607108	0.664441
	$\tau_{2_{opt}}(exp)$	1.951	2.530	2.975
	$\rho(\tau_{2_{opt}}(exp))$	0.492374	0.607155	0.664925
	$\omega_{2_{opt}}(exp)$	1.950	2.529	2.974
	$\rho(\omega_{2_{opt}}(exp))$	0.493127	0.607901	0.664657

Table 7 Computation of $\rho(\mathcal{H}(\tau_{1_{opt}}, \tau_{2_{opt}}, \omega_{2_{opt}}, a))$ for various values of the parameter a (Case 1, $p = 40$).

a	$\tau_{1_{opt}}$	$\tau_{2_{opt}} = \omega_{2_{opt}}$	$\rho(\mathcal{H}(\tau_{1_{opt}}, \tau_{2_{opt}}, \omega_{2_{opt}}, a))$
0	2.18851E-001	1.229935E-001	0.883807
10	2.18851E-001	5.515564E-002	0.883808
10^2	2.18851E-001	9.248083E-003	0.883808
10^3	2.18851E-001	9.919351E-004	0.883809
10^4	2.18851E-001	9.991876E-005	0.883810
10^5	2.18851E-001	9.999187E-006	0.883807
10^6	2.18851E-001	9.999919E-007	0.883807
10^7	2.18851E-001	9.999992E-008	0.883807
10^8	2.18851E-001	9.999999E-009	0.883815
10^9	2.18851E-001	1.000000E-009	0.886425
10^{10}	2.18851E-001	1.000000E-010	0.883879

Table 8 ITER, CPU and RES for the testing methods

		p	8	16	24	32	40	48
		n	64	256	576	1024	1600	2304
		m	128	512	1152	2048	3200	4608
		m+n	192	768	1728	3072	4800	6912
Case 1	GSOR	ITER	46	86	126	167	207	248
		CPU	0,05	0,36	3,71	22,47	86,47	258,28
		RES	6,79E-10	9,04E-10	9,79E-10	8,97E-10	9,74E-10	9,44E-10
	GMESOR	ITER	46	86	126	167	207	248
		CPU	0,05	0,36	3,71	22,61	86,78	258,93
		RES	6,79E-10	9,04E-10	9,79E-10	8,97E-10	9,74E-10	9,44E-10
	Simplified GMPSD	ITER	46	86	126	167	207	248
		CPU	0,05	0,35	3,71	22,59	86,54	258,28
		RES	7,03E-10	9,12E-10	9,83E-10	8,99E-10	9,75E-10	9,45E-10
	PHSS(a^*)	ITER	24	35	44	51	57	63
		CPU	0,34	5,24	34,52	147,95	472,87	1247,80
		RES	6,19E-10	9,62E-10	7,63E-10	7,36E-10	9,63E-10	8,82E-10
	GMRES	ITER	73	176	285	386	506	606
		CPU	0,33	9,24	155,67	1.240,62	6.352,04	22.214,42
	GMRES(100)	ITER	73	327	831	1794	3436	9965
		CPU	0,26	16,12	404,70	5.417,99	41.626,66	356.831,22
	PGMRES	ITER	76	143	207	275	344	410
		CPU	0,50	11,35	130,19	956,58	4.557,29	15.684,49
	PGMRES(100)	ITER	76	178	321	509	1038	1281
		CPU	0,36	11,02	172,80	1.615,68	12.838,90	46.595,50
Case 2	GSOR	ITER	65	124	182	241	300	359
		CPU	0,07	0,42	4,11	24,55	93,24	278,67
		RES	8,35E-10	8,25E-10	9,32E-10	9,14E-10	9,19E-10	9,35E-10
	GMESOR	ITER	65	124	182	241	300	359
		CPU	0,06	0,40	4,08	24,41	93,14	278,48
		RES	8,35E-10	8,25E-10	9,32E-10	9,14E-10	9,19E-10	9,35E-10
	Simplified GMPSD	ITER	65	124	182	241	300	359
		CPU	0,07	0,39	4,11	24,51	93,20	278,08
		RES	8,55E-10	8,30E-10	9,35E-10	9,15E-10	9,20E-10	9,35E-10
	PHSS(a^*)	ITER	29	43	53	62	69	76
		CPU	0,34	5,28	34,64	148,51	474,44	1261,96
		RES	9,88E-10	6,53E-10	7,99E-10	8,48E-10	9,61E-10	9,72E-10
	GMRES	ITER	73	176	285	386	506	606
		CPU	0,33	9,24	155,67	1.240,62	6.352,04	22.214,42
	GMRES(100)	ITER	73	327	831	1794	3436	9965
		CPU	0,26	16,12	404,70	5.417,99	41.626,66	356.831,22
	PGMRES	ITER	75	164	253	347	446	537
		CPU	0,50	12,18	151,43	1.168,31	5.761,78	20.131,99
	PGMRES(100)	ITER	75	275	544	997	*	*
		CPU	0,36	15,31	279,17	3.072,70	≫ 72h	≫ 72h

5 Remarks and Conclusions

In this paper we studied the impact of two different preconditioning matrices on the convergence of iterative methods for the solution of the augmented linear system (1) when the coefficient matrix $\mathcal{A}$ is of the form (2). We assumed that $A \in \mathbb{R}^{m \times m}$ was a symmetric positive definite matrix and $B \in \mathbb{R}^{m \times n}$ was a matrix of full column rank, where $m \geq n$, whereas Q was a symmetric positive or negative definite matrix. Under these assumptions we were able to find sufficient conditions for the GMESOR and GMPSD iterative methods to converge. Further, using a geometric analysis analogous to Varga [50] we determined the optimum values of the parameters of all methods studied such

as to attain the maximum rate of convergence. From our analysis it was shown that GMESOR and GMPSD are equivalent since they have the same spectral radius for the optimum values of their parameters, which is given by (93). This result was verified by our numerical experiments, where the simplified GMPSD, the GMESOR and the GSOR methods require approximately the same computing time. Moreover, all the aforementioned methods outperform the PHSS(a^*), GMRES, GMRES(#), PGMRES and PGMRES(#) methods considerably with respect to CPU times. It is worth mentioning that, for the saddle point problem, the GMPSD method has a similar behavior as the Modified PSD (MPSD) method for two-cyclic matrices [34]. Indeed, in [34] we proved the equivalence of MPSD and MSOR methods for two-cyclic matrices in case the eigenvalues of the Jacobi matrix are either all real or all imaginary. However, it is believed that this equivalence will not hold for the case where the eigenvalues of the J matrix are complex.

References

1. T. Arbogast and M. S. M Gomez, A discretization and multigrid solver for a Darcy-Stokes system of three dimentional vuggy porous media, Comput. Geosci, Vol. 13, pp. 331-348, 2009.
2. K. Arrow, L. Hurwicz and H. Uzawa, Studies in Nonlinear Programming, Stanford University Press, Stanford, 1958.
3. C. Bacuta and P. Monk, Multilevel discretization of symmetric saddle point systems without the discrete LBB condition, Appl. Numer. Math., Vol. 62, Issue 6, pp. 667-681, 2012.
4. C. Bacuta, L. Shu, Multilevel Gradient Uzawa Algorithms for Symmetric Saddle Point Problems, J. Sci. Comp., doi:10.1007/s10915-013-9697-9, 2013.
5. C. Bacuta, Cascadic Multilevel algorithms for saddle point systems, math. Num. Anal., arXiv:1305.2449v1, 2013.
6. Z. Z. Bai, G. H. Golub and M. K. Ng, Hermitian and skew-Hermitian splitting methods for non-Hermitian positive definite linear sytems, SIAM J. Matrix Anal. Appl., Vol. 24, pp. 603-626, 2003.
7. Z. Z. Bai, G. H. Golub and J.-Y. Pan, Preconditioned Hermitian and skew-Hermitian positive semidefinite linear systems, Numer. Math., Vol. 98, pp. 1-32, 2004.
8. Z. Z. Bai, B. N. Parlett and Z.-Q. Wang, On generalized succesive overrelaxation methods for augmented linear systems, Numer. Math., Vol. 102, pp. 1-38, 2005.
9. Z. Z. Bai, G. H. Golub and C.-K. Li, Optimal parameter in Hermitian and skew-Hermitian splitting method for certain two-by-two block matrices, SIAM J. Sci. Comput., Vol. 28, pp. 583-603, 2006.
10. Z. Z. Bai and G. H. Golub, Accelerated Hermitian and skew-Hermitian splitting iteration methods for saddle-point problems, IMA J. Numer. Anal., Vol. 27, pp. 1-23, 2007.
11. Z. Z. Bai, G. H. Golub and C.-K. Li, Convergence properties of preconditioned Hermitian and skew-Hermitian splitting methods for non-Hermitian positive semidefinite matrices, Math. Comput., Vol. 76, pp. 287-298, 2007.
12. Z.-Z. Bai and Z.-Q. Wang, On parameterized inexact Uzawa methods for generalized saddle point problems, Lin. Alg. and its Appl., Vol. 428, pp. 2900-2932, 2008.

13. M. Benzi, G. H. Golub and J. Liesen, Numerical solution of saddle point problems, Acta Numerica, pp. 1-137, 2005.
14. J. H. Bramble, J. E. Pasciak and A. T. Vassilev, Analysis of the inexact Uzawa algorithm for saddle point problems, SIAM J. Numer. Anal., Vol. 34, pp. 1072-1092, 1977.
15. W. L. Briggs , V. E Henson and S. F. McCormick, A Multigrid Tutorial, Second edition edn., SIAM, 2000.
16. H. C. Elman and G. H. Golub, Inexact and preconditioned Uzawa algorithms for saddle point problems, SIAM J. Numer. Anal., Vol. 31, pp. 1645-1661, 1994.
17. H. C. Elman and M. H. Schultz, Preconditioning by fast direct methods for nonself-adjoint nonseparable elliptic equations, SIAM J. Numer. Anal., Vol. 23, pp. 44-57, 1986.
18. H. C. Elman, D. J. Silvester and A. J. Wathen, Performance and analysis of saddle point preconditioners for the discrete steady-state Navier-Stokes equations, Numer. Math., Vol. 90, pp. 665-688, 2002.
19. D. J. Evans and N. M. Missirlis, The preconditioned simultaneous displacement method (PSD method) for elliptic difference equations, Mathematics and Computers in Simulation, Vol. 22, pp. 256-263, 1980.
20. B. Fischer, R. Ramage, D. J. Silvester, A. J. Wathen, Minimum residual methods for augmented systems, BIT, Vol. 38, pp. 527-543, 1998.
21. G. H. Golub, X. Wu and J.-Y Yuan, SOR-like methods for augmented systems, BIT, Vol. 41, pp. 71-85, 2001.
22. F. J. Gaspar, F. J. Lisbona, C. W. Oosterlee and P. Vabishchevich, An efficient multigrid solver for a reformulated version of the poroelasticity system, Comput. Methods Appl. mech. Engrg, Vol. 196, pp. 1447-1457, 2007.
23. A. Gordon and C. Powell, On solving stochastic collocation systems with algebraic multigrid, IMA Journal of Numerical Analysis, Vol. 32, pp. 1051-1070, 2012.
24. M. Griebel, T. Dornseifer and T. Neunhoffer, Numerical simulation in fluid dynamics, a practical introduction, SIAM, Philadelphia, 1998.
25. M. Griebel, D. Oeltz and M. A. Schweitzer, An algebraic multigrid method for linear elasticity, SIAM J. Sci. Computing, Vol. 25, Issue 2, pp. 385-407, 2003.
26. M. Griebel, B. Metsch, D. Oeltz and M. A. Schweitzer, Coarse grid classification: A parallel coarsening scheme for algebraic multigrid methods, Num. Lin. Algebra Applic., Vol. 13, Issue (2-3), pp. 193-214, 2006.
27. M. Griebel, B. Metsch and M. A. Schweitzer, Coarse grid classification - Part II: Automatic coarse grid agglomeration for parallel AMG, Preprint 271, Sonderforschungsbereich 611, Universität Bonn, 2006.
28. M. Griebel, B. Metsch and M. A. Schweitzer. Coarse Grid Classification: AMG on Parallel Computers. In Gernot Müunster, Dietrich Wolf, and Manfred Kremer, editors, NIC Symposium 2008, Vol. 39 of NIC Series, pp. 299-306, 2008.
29. A. Hadjidimos, Accelerated Overrelaxation Methods, Math. of Comput., Vol. 149-157, 1978.
30. S. Hamilton, M. Benzi and E. Haber, New multigrid smoothers for the Oseen problem, Num. Linear Algebra Applic., Vol. 17, pp. 557-576, 2010.
31. M. R. Hestenes and E. Stiefel, Methods of Conjugate Gradients for Solving Linear Systems, Journal of Research of the National Bureau of Standards, Vol. 49, Issue 6, pp. 409-436, 1952.
32. C. Lu, X. Jiao and N. M. Missirlis, A hybrid geometric + algebraic multigrid method with semi-iterative smoothers, Num. Linear Algebra Applic., Vol. 21, Issue 2, pp. 221-238, 2014.
33. C. J. Li, Z. Li, D. J. Evans and T. Zhang, A note on an SOR-like method for augmented systems, IMA J. Numer. Anal., Vol. 23, pp. 581-592, 2003.
34. M. A. Louka, N. M. Missirlis and F. I. Tzaferis, Is modified PSD equivalent to modified SOR for two-cyclic matrices? Lin. Alg. and its Appl., Vol. 432, Issue 11, pp. 2798-2815, 2010.
35. B. Metsch, Algebraic Multigrid (AMG) for Saddle Point Systems, Ph.D thesis, Mathematisch Naturwissenschaftlichen Fakultät, Universität Bonn, 2013.
36. N. M. Missirlis and D. J. Evans, On the convergence of some generalised preconditioned iterative methods, SIAM J. Numer. Anal., Vol. 18, pp. 591-596, 1981.
37. N. M. Missirlis, Convergence Theory of Extrapolated Iterative Methods for a Certain Class of Non-Symmetric Linear Systems, Numer. Math., Vol. 45, pp. 447-458, 1984.

38. N. M. Missirlis and D. J. Evans, The modified preconditioned simultaneous displacement (MPSD) method, Mathematics and Computers in Simulation, Vol. 26, pp. 257-262, 1984.

39. M. F. Murphy, G. H. Golub and A. J. Wathen, A note on preconditioning for indefinite linear systems, SIAM J. Sci. Comput., Vol. 21, pp. 1969-1972, 2000.

40. C. W. Oosterlee, F. J. Gaspar, Multigrid relaxation methods for systems of saddle point type, Appl. Num. Math., Vol. 58, pp. 1933-1950, 2008.

41. Y. Saad and M. H. Schultz, GMRES: A Generalized Minimal Residual algorithm for solving nonsymmetric linear systems, SIAM J. Sci. Stat. Comput., Vol. 7, No. 3, pp. 856-869, 1986.

42. Y. Saad, Iterative Methods for Sparse Linear Systems, Second edition, SIAM, Philadelphia, 2003.

43. V. Schulz and G. Wittum, Transforming smoothers for PDE constraint optimization problems, Comput. Vis. Sci, Vol. 11, Issue 4-6, pp. 207-219, 2008.

44. R. Simon and W. Zulehner, On Schwarz-type smoothers for saddle point problems with applications to PDE-constrained optimization problems, Numer. Math., Vol. 111, pp. 445-468, 2009.

45. S. Takacs and W. Zulehner, Multigrid methods for Elliptic Optimal Control Problems with Neumann Boundary Conditions, Proceedings of ENUMATH 2009 the 8th European Conference on Numerical Mathematics and Advanced Applications, Uppsalla, pp. 855-863, 2009.

46. S. Takacs and W. Zulehner, Convergence analysis of multigrid methods with collective point smoothers for optimal control prolems, Comput. Visual Sci., Vol. 14, pp. 131-141, 2011.

47. P. J. Taylor, A generalization of systematic relaxation methods for consistently ordered matrices, Numer. Math., Vol. 13, pp. 377-395, 1969.

48. U. Trottenberg, C. W. Oosterlee and A. Schuller, Multigrid, Academic Press, 2000.

49. H. A. Van der Vorst, Iterative Krylov Methods for Large Linear Systems, Cambridge University Press, Vol. 13, 2003.

50. R. S. Varga, Matrix Iterative Analysis, Prentice-Hall, Inc. Englewood Cliffs, N.J., 1962.

51. J. Volker, On the efficiency of linearization schemes and coupled multigrid methods in a simulation of a 3D flow around a cylinder, Int. J. Numer. Meth. Fluids, Vol. 50, pp. 845-862, 2006.

52. C. L. Wang and Z. Z. Bai, Sufficient conditions for the convergent splittings of non-Hermitian positive definite matrices, Linear Algebra Appl., Vol. 330, pp. 215-218, 2001.

53. M. Wang and L. Chen, Multigrid methods for the Stokes Equations using Distributive Gauss-Seidel Relaxation based on the Least Squares Commutator, J. Sci. Comp., Vol 56, pp. 409-431, 2013.

54. A. J. Wathen and D. J. Silvester, Fast iterative solution of stabilized Stokes systems. Part I: Using simple diagonal preconditioners, SIAM J. Numer. Anal., Vol. 30, pp. 630-649, 1993.

55. D. M. Young, Iterative Solution of Large Linear Systems, Academic Press, New York, 1971.